SECOND EDITION

College Physical Science

Wendell H. Slabaugh and Alfred B. Butler

Phy
4210
17.5.88
8734
8734
STATE INSTITUTE OF EDUCATION
Govt. of West Bengal

College Physical Science

8734
STATE INSTITUTE OF EDUCATION
Govt. of West Bengal

College Physical Science

SECOND EDITION

Wendell H. Slabaugh
Oregon State University

Alfred B. Butler
Washington State University

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road

PRENTICE-HALL, INC.
Englewood Cliffs, New Jersey

PRENTICE-HALL INTERNATIONAL, INC., *London*
PRENTICE-HALL OF AUSTRALIA, PTY., LTD., *Sydney*
PRENTICE-HALL OF CANADA, LTD., *Toronto*
PRENTICE-HALL OF INDIA (PRIVATE) LTD., *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*

S.C.E.R.T., West Bengal
Date .....
Acc. No. 4218

COLLEGE PHYSICAL SCIENCE
Second Edition
Wendell H. Slabaugh and Alfred B. Butler

© 1965 by PRENTICE-HALL, INC., *Englewood Cliffs, New Jersey*

All rights reserved. No part of this book may be reproduced in any form, by mimeograph or any other means, without permission in writing from the publisher.

Library of Congress Catalog Card Number
65 – 19169

Printed in the United States of America

14735 C

Preface

This is a textbook for courses in physical science at the college level. It is designed so that it may be used for either a shorter one semester course or for a full year's course. That it may be used for the shorter course is evidenced by the fact that the chapters on heat energy constitute a unit that has been successfully used for several years by one of the authors. Within the book are other unit groupings that could also be selected as the basis for shorter courses. It is assumed that the student has a typical, non-scientific background; the topics are developed with a minimum of mathematical treatment.

The various disciplines of physical science have been developed for the most part by specialists in those disciplines. The typical physicist claims little or no professional interest in chemistry, while the typical astronomer cares little for geology. However, there is a growing trend to bridge over the gaps between the different areas—in such fields as geochemistry, astrophysics, physical chemistry, geophysics, and so on—and this trend confirms our idea that the physical sciences together constitute one major field of human knowledge. In spite of the profitable results that are presently coming from these bridge areas, there are still many areas within the parent disciplines that defy integration into a common subject. The authors believe, nevertheless, that for the non-science student the advantages of studying physical science as an integrated subject warrant further efforts to correlate and integrate the various areas of physical science into a comprehensible whole.

As examples of the benefits to be derived from such an integration, attention is called to Chapters 3 and 4 where stream erosion, downslope movements, and tides are discussed in terms of gravitation and the inclined plane. The classical topics of force and motion are correlated with diastrophism, earthquakes, theories of mountain formation, and the solar system in Chapters 5, 6, and 7. And electricity, magnetism, and wave motion form the background to the material on atomic structure

and science in Chapters 8, 10, 12, and 13. As a result of this treatment of subject matter, it is difficult to count the number of chapters devoted respectively to physics, chemistry, astronomy, or geology. Part of no less than eight chapters deal with geology, astronomical topics are found in at least ten chapters, and so on.

Above all, the authors believe that for the non-science student an understanding of the four primary concepts of space, mass, time, and energy is crucial to an introductory study of physical science. Consequently these four topics are used as the basis for the development of the many scientific principles taken up in the book. Topics dealing with energy, for instance, appear in almost every chapter. Another innovation reflects the opinion of the authors that physical science should not be considered in a vacuum. For this reason, a concluding chapter concerning some of the philosophical aspects of physical science is designed to assist the student in orienting himself in a world in which science is a, if not the, dominating influence. In short, an attempt has been made not only to integrate subject matter among the various fields, but to associate these fields with broader elements in our modern life. It is hoped that such an approach accenting the wider implications of science will provide a more meaningful experience for the student in his study of physical science.

A comment about the brief historical sketches that introduce certain topics. Although the treatment is not intended to be primarily historical, the authors think that the inclusion of a certain amount of historical background accomplishes two valuable objectives. first, it makes the student aware of the chronological development of scientific principles and of the human effort that must be expended before such principles are verified, and second, of equal importance, it presents the subject matter in a context that the student can relate to his other historical and humanistic learning.

Where possible, a narrative style of presentation is used in order to stimulate student interest. Illustrations are used freely to amplify the discussions, and, in order to help the student sift the grain from the chaff, exercises at the end of each chapter point up those topics and concepts that are considered fundamental. Enough exercises are provided to enable the instructor to use them in an individual and flexible way.

Acknowledgments. The authors are deeply grateful to all their colleagues, both present and past, with whom many of the ideas in this book have been discussed, and in particular to Professor A. B. Ignasy and others at Kansas State University where the seeds of the integrated approach embodied in this text were germinated. The indispensable help of Herbert Weiss in editing the book is gratefully acknowledged.

[illegible]
[illegible]

Contents

ONE
The Nature of Physical Science 1

Physical Science Defined. The Emergence of Science. The Branches of Physical Science. The Scientific Method. Physical Science as a Profession. Summary.

TWO
Space, Mass, Time, and Energy 11

Numerical Abbreviations. Length and Distance. Area, Volume, and Space. Mass and Weight. The Analytical Balance. The Metric System of Mass and Volume. Numerical Abbreviations. Density. Time. Time from the Sun and Stars. Time from the Rocks. Time and Man. What Is Energy? Summary.

THREE
Gravity 31

Speed and Velocity. Scalars and Vectors. Acceleration. The Inclined Plane. Gravity and Falling Bodies. The Acceleration Due to Gravity. Mathematical Formulas. Newton's Law of Gravitation. Mass and Weight. Summary.

FOUR
Gravity in Natural Processes 45

Gravity as a Force in Erosion. Erosion and the Landscape. Formation of Sedimentary Rocks. Glaciers. Landslides. Variations in Gravity. Inertia and Archimedes' Principle. The Tides. Gravity and the Atmosphere. Summary.

FIVE
Force and Motion 59

Newton's First Law. Newton's Second Law. Momentum. Work. Newton's Third Law. Combinations of Forces. Forces in Equilibrium. Unbalanced Forces. Circular Motion and Centripetal Force. Kinetic and Potential Energy. Conservation of Energy. Summary.

SIX
Force and Motion in Natural Processes **96**

Warps. Folds. Geosynclines. Faults and Joints. Earthquakes. The Earth's Structure from Earthquake Studies. Metamorphism. The Origin of Mountains. Theories of Mountain Formation. Summary.

SEVEN
The Solar System **115**

Early Greek Concepts of the Solar System. The Ptolemaic Universe. Copernicus and the Heliocentric Theory. Theory in Physical Science. The Planets and Their Orbits. Kepler's Laws. The Sun. The Earth. Moons and Satellites. Planetoids, Comets, and Meteors. The Origin of the Solar System. Summary.

EIGHT
Electrostatics and the Electron **141**

Electrostatics. The Nature of Positive and Negative Electricity. The Electroscope. Insulators and Conductors. Electrostatic Attraction and Repulsion: Coulomb's Law. The Electron. Radioactivity. Protons and Neutrons. Induction. Electrostatics in Nature: Lightning and Aurora Borealis. Summary.

NINE
Electricity and Magnetism **159**

Electricity. Electrical Units. Electrical Energy and Power Magnetism. The Earth's Magnetic Field. Magnetism from Electricity. Electricity from Magnetism. Mechanical Energy from Electrical Energy. Direct and Alternating Currents. Series and Parallel Circuits. Summary.

TEN
Wave Motion **178**

Wave Motion. Transverse, Longitudinal, and Complex Waves. Wave Length and Velocity. Source of Sound Waves. Sources of Light Waves. Reflection of Waves. Refraction of Waves. Interference of Waves. Optical Instruments. Color. Rainbows. X Rays. Wave Motion in Natural Processes. Summary.

ELEVEN
Atomic Theory **202**

Historical Theories of Atoms. The Chemical Revolution. Dalton's Atomic Theory. Atoms and Molecules. Modern Atomic Theory. The Mass Spectrograph. The Law of Combining Volumes. Atomic Weights. Chemical Equations. Atoms and X-ray Diffraction. Summary.

TWELVE
Atomic Structure **224**

X Rays and Radioactivity. Inside the Atom. The Simplest Atom. The Hydrogen Spectrum and the Photoelectric Effect. The Bohr Atom. More Complex Atoms. Atomic Numbers. Summary.

THIRTEEN
The Elements **243**

The Discovery of the Elements. The Reactivity Series of the Elements. The Natural Occurrence of the Elements. The Abundance of the Elements. The Periodic Classification of the Elements. Uses of the Periodic Table. Metals and Non-metals. The Periodic Table and Atomic Structure. Summary.

FOURTEEN
Chemical Bonds **262**

Atomic Radii. Valence and Electrolysis. Ionization Potential and Electron Affinity. Covalence. Coordinate Valence. Other Types of Chemical Bonds. Oxidation and Reduction. Summary.

FIFTEEN
Ionic Substances **284**

Electrolytes. Solutions of Ionic Substances. Oxides. Acids. Bases. Neutralization of Acids and Bases. Salts. The pH of Water. Hydrolysis. Water in Natural Processes. Erosion of Limestone. Erosion of Granite and Other Rocks. Summary.

SIXTEEN
Covalent Substances **307**

Elementary Covalent Substances. Organic and Inorganic Substances. Silicon Compounds. Structure in Carbon Compounds. Hydrocarbons. Polymers. Cyclic Hydrocarbons. Alcohols, Acids, and Esters. Amines and Amides. Summary.

SEVENTEEN
Chemical Reactions **331**

Kinds of Energy. Exothermal and Endothermal Reactions. The Mechanism of Chemical Reactions. Activation Energy and Catalysts. Rate of Chemical Reaction. Chemical Equilibrium. Factors that Affect Chemical Equilibrium. The Ammonia Synthesis. Reaction Mechanisms in Natural Processes. Summary.

EIGHTEEN
Biophysical Processes **351**

Autotrophic and Heterotrophic Organisms. Carbohydrates. Lipids. Proteins and Amino Acids. Fermentation. Respiration. Photosynthesis. Minerals and Vitamins. Chemotherapy. Fossilization of Carbohydrates. Summary.

NINETEEN
Electrochemical Energy 372

Electrochemical Energy. Electrochemical Cells. Electrolysis. Electrolytic Production of Elements. Corrosion. Semiconductors. Summary.

TWENTY
Heat Energy 387

The Measurement of Temperature. Heat and Temperature. Specific Heat. Heat Energy Transformed. Expansion Due to Heat. Methods of Heat Transfer. Summary.

TWENTY-ONE
The Kinetic Theory of Matter 407

Postulates of the Kinetic Theory. Intermolecular Forces. Surface Tension. Capillarity. Osmosis. Behavior of Gases. Velocity of Gas Molecules. Probable Distribution of Molecular Phenomena. Brownian Movement. Diffusion. Colloids. Review of the Kinetic-Molecular Theory. Summary.

TWENTY-TWO
Solids, Liquids, and Gases 429

Solids. Liquids. Gases. Relationship between the Solid and Liquid States. Heat of Fusion. Heat of Vaporization. Summary of Heat Exchange Problems. Effect of Pressure on the Boiling Point. Effect of Pressure on the Freezing Point. Sublimation. Critical Point. Pressure, Temperature, and Volume Relationships. Summary.

TWENTY-THREE
Heat in Natural Processes: I 451

Our Star—the Sun. Temperature and Life in the Solar System. The Greenhouse Effect. The Greenhouse Effect and the Glacial Ages. Water in the Air. The Measurement of Relative Humidity. Condensation in Rising Air Masses. Air Masses: Their Source and Movement. Cold Fronts. Warm Fronts. Circulation of the Atmosphere and the Coriolis Effect. Circulation of the Oceans. Circulation of the Atlantic Ocean. Oceans and the Weather. Summary.

TWENTY-FOUR
Heat in Natural Processes: II 478

Sources of the Earth's Heat. Magma and Molten Rock. Intrusive Magmatic Activity. Chronology of Geologic Events. Types of Volcanic or Extrusive Activity. A Cycle of Volcanic Activity. Types of Ejected Magmatic Material. The Earth's Heat and Human Affairs. Phenomena in Lava Flows. Volcanic Activity in Continental United States. Hot Springs and Geysers. Heat in Biological Processes. Summary.

TWENTY-FIVE
Radioactivity **503**

The Nuclear Atom. Natural Radioactivity. Nuclear Stability. Artificial Transmutation. The Acceleration of Charged Particles. The Measurement of Radioactivity. Radioactivity in Tracer Studies. The Measurement of Time by Radioactivity. Summary.

TWENTY-SIX
Nuclear Energy **523**

The Source of Nuclear Energy. The Neutron. Nuclear Fission. The Utilization of Nuclear Energy. New Elements. Nuclear Fusion. Separation of Isotopes. Solar Energy. Summary.

TWENTY-SEVEN
The Universe and Cosmogony **542**

Constellations. The Nearer Stars. Variable Stars. Novae and Supernovae. Our Galaxy. Galaxies and the Universe. Evolution of the Universe. Summary.

TWENTY-EIGHT
The Availability of Energy **559**

Sources of Energy. Conservation of Mechanical Energy. Conservation of Energy in Atomic Systems. Conservation of Heat Energy. Heat Engines. The Efficiency of Heat Engines. The Availability of Useful Energy from Heat. Energy in Natural Processes. The Increasing Demands for Our Energy Resources. Summary.

TWENTY-NINE
New Horizons in Physical Science **577**

Modern Concepts of Space. Relativity. New Horizons in Research.

THIRTY
The Philosophy of Physical Science **592**

Observation in Physical Science. Cause and Effect. Probability and Hypothesis in Physical Science. Scientific Law. Social Attitudes toward Physical Science. The Social Responsibility of Scientists.

Index **603**

One

The Nature of Physical Science

We are born into a scientific world, and it is impossible to grow up today without coming into close contact with much of what science offers. The typical baby is ushered into the world amidst the gadgets and paraphernalia of a modern hospital. He is given the best chance for life through the use of efficient medicants, diet, and care. He grows up faster and stronger and healthier than any of his ancestors, and he will probably live many years longer. Much of his happiness and well-being stem directly from science. He wears materials and eats foods that were unknown to his grandparents, and he complacently views high-compression engines, plastics, supersonic jets, rockets, radar-guided missiles, and nuclear weapons as commonplace items that appear regularly in the newspapers.

He may even tend to take all these trappings for granted. He may not be concerned about our diminishing reserves of copper, petroleum, iron, or sulfur because science has so benevolently and adequately met all our previous shortages by supplying substitutes that are usually much better than the original. For example, not many years ago it was feared that a rapidly increasing population would eat up more than it could produce. Not only food but the natural supply of fertilizers would be used up and the human race would be decimated. But in the meantime man has learned how to make efficient fertilizers from water and air, which are, of course, an unlimited source of raw materials. Now, in this country, our food problems center on what we should do with surpluses—a far cry from the fear of famine.

In other words, we grow up in a world where science is a normal part of our lives, and we usually give it little

thought. But this in no way eliminates or detracts from its importance. If for some reason we had to go through a brief period without the items that science makes possible, we would soon appreciate the value of science in our lives. Not only would we be without enough food, but the water we drink and even the air we breathe might be poisonous. There would be no radio, television, or telephone. Our clothes would be drab and colorless, and most of us would have no transportation. For most people there would be no jobs, because modern industries, based on the technology nurtured by science, could not exist. There would be little if any spare time for recreation or the lighter pursuits because we would be fully occupied with major problems of survival. Even such creative activities as art, music, and drama would be reduced to a minimum because no one would have time to appreciate them. The world would, indeed, be a harsh and dreary place.

Science, fortunately, is with us, and as our civilization continues to grow and mature, we hope that not only science but all the other constructive activities of mankind—politics, economics, philosophy, and the arts—will progress at a corresponding rate. Science has succeeded so well in solving its many problems that men in other areas have turned to science as a means of solving what were once perhaps considered as unscientific problems. More and more the techniques and methods of science are being applied to these other areas.

Why study physical science? For one thing, new knowledge, whether or not it is of any practical use, is a great source of pleasure. And men are beginning to realize that science is the greatest adventure of our age and that its discoveries help satisfy man's search for beauty and understanding. Moreover, modern life demands that we have some appreciation of the role of science in our civilization. Men and women in every profession or vocation—engineering, medicine, farming, or running a home—profit from a knowledge of physical science. Finally, learning more about the world in which we live gives us a better understanding of how we, as human beings, fit into the scheme of things.

1-1 Physical Science Defined

Physical science is knowledge of the physical world—of the composition, structure, origin, and behavior of the materials that make it up. But science is more than knowledge; it is also the process by which that knowledge is obtained. This process, known as the *scientific method*, involves making observations under controlled conditions with an attitude of impartiality and openmindedness. With this tool at their command, scientists have not only accumulated a vast store of information, but have also succeeded in summarizing and integrating their information into natural laws and theories.

The discovery, collection, and classification of facts is known as *descriptive science.* The formulation of theories and laws that explain and unify these facts is known as *theoretical science.* In this book we shall see how descriptive and theoretical science are fused together to form the framework of physical science.

1-2 The Emergence of Science

The growth of science from its humble beginnings to the present may be divided into six periods. No sharp distinction exists between these periods, for one gradually evolved into another. Nor should we look upon the present status of science as the ultimate of achievement. Future generations will likely look back upon our age of science in much the same way as we now look back upon the earlier periods of science, because science will undoubtedly continue to grow and develop with our expanding civilization.

1. *Prehistory.* When and where did science begin? The time and place when man's curiosity about nature was transformed into a science is difficult to determine. The making of fire, the invention of the wheel, the accidental discovery of glass and pottery, and the fashioning of simple handtools all resulted from man's attempt to solve the innumerable problems of everyday life. When and how these discoveries were made we do not know, but it is known that as far back as five thousand years ago man was using numerical notations and mathematics, and making vinegar, soap, and medicinals of various kinds. He observed the sun and stars from season to season and invented calendars. He learned to cook food and clothe his body with furs and later with textiles. He learned to tint glass, dye textiles, and separate metals from their ores. He gradually mastered at least part of his environment—an achievement that distinctly separated him from the animals that roamed the forest and plain. The Biblical account of Noah's experiences even indicates that the manufacture of intoxicating beverages preceded some of the events in Genesis.

Although there may be some reluctance to view these early activities as scientific—many of them were crude and rudimentary when judged by our modern standards—one of the deepest of human traits, curiosity, impelled man to solve his problems, and this quality is still the mainspring of modern science. If necessity is the mother of invention, curiosity is certainly the mother of science. Since early men were motivated by the same urge as modern scientists, it is only fair to consider these early activities as the dawn of science.

2. *The Mediterranean Period (Egypt and Greece).* In this period, from the earliest recorded history to the fall of Rome, science made some remarkable progress. The universe was considered as an ordered system, and there were serious attempts to theorize and explain the structure and behavior of

matter. The early concepts of astronomy, mathematics, and chemistry were probably developed in Egypt, where an alphabet, writing, and papyrus scrolls were also invented. Architecture and engineering were advanced to a high degree, as attested by the construction of the pyramids.

The scholarly Greeks were quick to expand on the numerical abstractions invented by the Egyptians and Babylonians. Pythagoras taught in the sixth century B.C. that numerical relationships can be found in all nature. Empedocles (490–430 B.C.) announced that the world is composed of four elements—earth, air, fire, and water. Leucippus and Democritus proposed that all matter is composed of indestructible tiny particles called atoms, and Euclid in about 300 B.C. developed the system of geometry that is used today. Archimedes experimented with the lever and with the principle of specific gravity, and through them made fundamental discoveries in mechanics and hydrostatics. During the early years of the Christian Era, science pushed forward chiefly in mathematics and astronomy. Pliny the Elder collected a large number of books containing the scientific knowledge that had been accumulated up to the end of the first century A.D. Ptolemy of Alexandria, who lived around A.D. 150, explained the motion of the planets in a theory that was not seriously questioned for 1400 years.

3. *The Middle Ages.* After the decline of the Mediterranean civilizations, scientific inquiry virtually came to a stop. In its place arose the practice of mystic arts and sorcery. Men were driven by a strange urge to seek methods of making gold from iron and lead. It was during this time that the *alchemists* tried to find the philosopher's stone—a substance that would give universal wisdom to its owner and also convert ordinary metals to gold. But the contribution of alchemy to science was of negligible value; the alchemist left little or no record of his experiments and made no attempt to correlate theory with experiment. Since men were more interested in religion and other pursuits than in science, they failed to develop even the simplest instruments to test or advance the theories of earlier science. Communication among the feudal manors was poor, and both travel and publication were almost nonexistent. Only a few names in science come down to us from the Middle Ages. Roger Bacon stressed the experimental method in science in the thirteenth century, and Paracelsus showed the value of chemistry in producing medicines.

4. *The Renaissance.* After centuries of indifference to scientific investigation, in the fifteenth century men began to look at nature not as the product of a fearsome, inscrutable God, but as something to probe and understand. Men became intrigued with the mysteries of nature underfoot and overhead, especially with those in the night sky. During this period of rebirth that we call the Renaissance, Nikolaus Copernicus (1473–1543) studied the motion of the planets and discovered that the sun, rather than the earth, is the

center of the solar system. Galileo Galilei, a century later, confirmed the findings of Copernicus. Against the dogmas and authority of the Church, he championed the validity of the blasphemous theory. Although the Church made him renounce his theory, his experiments marked the beginning of the end of medieval secrecy, superstition, and prejudice.

Out of the inquiring spirit of the Renaissance grew the methods and early knowledge of what we now call physical science. Alchemy gave way to chemistry. Physics gained a firm basis by applying mathematics to the natural processes of gravity, motion, and force. Astronomy advanced as more and more powerful telescopes and more accurate measuring instruments were invented to observe and correlate the motions of the heavenly bodies. Important scientists of this period, in addition to Copernicus and Galileo, were the astronomers, Tycho Brahe and Johannes Kepler, and the physicist, Robert Boyle, who discovered several properties of gases that are basic to modern physics and chemistry.

5. *The Growth of Science.* At the beginning of the seventeenth century, physical science gradually entered its modern age. Many mistakes were made and much groping in the darkness took place (but when don't they?). Science, however, strode forward and produced developments, the effects of which are still evident in our present-day scientific thinking. Isaac Newton advanced his famous analysis of gravitation and his laws of motion. Huygens and Newton studied light and lenses and discovered laws and relationships that are valid today. Organic and biological chemistry made the first accurate description of red blood corpuscles. The stimulating, though erroneous, phlogiston theory was proposed to explain the process of combustion.

During the eighteenth century, Benjamin Franklin performed his well-known experiments on electricity, Watt invented the steam engine, and the analytical balance was introduced into chemical analysis. Antoine Lavoisier, Joseph Priestly, Henry Cavendish, and a host of others pursued the experimental study of physics and chemistry, giving us the basis for the laws of the conservation of energy and the conservation of mass.

The nineteenth century was probably the golden age of modern science. Many outstanding and significant advances were made, particularly in formulating and discovering the fundamentals of physical science. The concept of atoms was modernized and accommodated to fit the problems of chemistry by John Dalton. Many new chemical elements were discovered by Davy, Berzelius, Bunsen, and others. Michael Faraday and Joseph Henry added greatly to the knowledge of electricity and magnetism. The last century was also a time of great growth in scientific theory. Most of our modern theories and concepts in science were devised during the latter half of that century. Indeed, it has been said that the most revolutionary thinking in science occurred during the ten years 1895 to 1905. In that short period X rays

were first produced, radioactivity was discovered, the quantum concept of light was developed by Max Planck, atomic disintegration was demonstrated by Rutherford and Soddy, and Albert Einstein proposed his famous theory of relativity.

6. *The Present.* Physical science today is largely an outgrowth of the ideas and concepts that were introduced in earlier years. The frontiers have been pushed outward by improving the theories and filling in the many gaps with experimental data. Several advances merit a listing here, though all of them will be discussed in greater detail in later chapters of this book.

A direct application of the Einstein theory was made in atomic or nuclear energy, and the predictions he made years ago have been verified experimentally in the atomic and H-bombs. The properties of light and the broader aspects of wave motion have provided a key to the structure of the atom, and the quantum theory has opened up the innermost secrets of the atom's interior. In chemistry hundreds of thousands of new substances have been synthesized, leading to new drugs, plastics, synthetic fibers and rubbers, and new metals and alloys. And, aided by the fundamental ideas of physics and chemistry, astronomy and geology have literally reached out into new worlds.

1-3 The Branches of Physical Science

Each of the four sciences comprising physical science—physics, chemistry, geology, and astronomy—developed as a separate field of study. As these fields have grown, however, they have overlapped with the result that today it is difficult to tell where one ends and the others begin. Although not exactly a branch of physical science, mathematics is an indispensable tool that serves them all.

Physics is the study of matter and motion. Physicists investigate heat, light, mechanics, electricity, sound, radiations, and the structure of matter, and they have supplied the major share of the basic ideas and concepts in physical science. Other sciences have profited greatly from the application of this knowledge acquired in physics.

Chemistry is the science of substances—their composition, their structure, and the reactions that change them into other substances. Chemical processes are at the core of the manufacture of many modern products: dyes, synthetic rubber, paints, synthetic fibers, plastics, refrigerants, insecticides, drugs, medicines, and many others.

Geology is the study of the history of the earth and the forces that change its shape—from earthquakes to tiny streams. A geologist is concerned with the classification and identification of rocks and minerals, the formation and location of natural resources, the causes of weather, the currents of the ocean—in short, with all the natural forces acting on the earth's surface, subsurface, and atmosphere.

Astronomy is the study of the celestial bodies. We are indebted to the astronomer for our present knowledge of the size, shape, and character of the solar system, the characteristics of the stars, and the structure of the universe.

Essential to physical science is *mathematics*—the abstract treatment of quantities. Scientists are continually asking: How much? How long? How many? How far? Lord Kelvin, one of England's most famous scientists of the nineteenth century, once observed: "When you can measure what you are speaking about and express it in numbers, you know something about it." Measurement is the keystone of the scientific method. But after the scientist has done all his measuring, he must combine his figures to arrive at a conclusion, and that process of working with numbers in the abstract is mathematics. Without mathematics, science could never have become more than mere curiosity, and for this reason mathematics is often called the "queen of sciences."

Each branch of physical science is divided into specialized subdivisions. Physics embraces mechanics, optics, electronics, thermodynamics, acoustics, quantum mechanics, statistics, and atomic structure. Chemistry includes analytical, inorganic, organic, biological, and theoretical chemistry. Special divisions in geology are physical, historical, and stratigraphical geology. But as we have seen, the branches of physical science often have merged. This meshing of disciplines has created many fields of specialization that are concerned with two or more sciences—for example, geophysics, geochemistry, physical chemistry, chemical physics, and astrochemistry. As physical science matures, the identity of its components is becoming less and less distinct. Moreover, some of the physical sciences are uniting with the biological sciences in such areas as biophysics and biochemistry. Great opportunities in science now lie in the application of the knowledge from one field to the problems in another field; some of the most recent and significant advances in science have resulted from cooperative studies between fields.

For example, some of the new antibiotics are the result of the successful combination of knowledge from several branches of science. Penicillin, a popular antibiotic, was discovered by the English microbiologist, Alexander Fleming. He was assisted by chemists who helped formulate the nutritive culture in which the special molds were grown. From these molds the biologically active component, penicillin, was separated, and its actual molecular structure was found to be so complex that both chemical and physical methods were required for its identification. Even the most powerful tools of the physicist and the physical chemist were needed. Organic chemists then remade the molecule in order to fully verify its structure. As to the clinical use of penicillin, the biochemist and the medicinal chemist studied its biological activity and determined the ways it could be employed most effectively.

Atomic energy, new metals and alloys, synthetic rubber

and fibers, and television are additional examples of the results of such cooperative efforts, the success of which indicates that the branches of physical science should be considered as members of one body. And beyond that, the fact that many scientific activities now range across many different sciences should make us realize that these divisions are only arbitrary and that all science should be treated as a broad, interrelated field of human knowledge. With this thought in mind we shall explore physical science, not as a collection of unrelated phenomena, but as an integrated and unified approach to understanding the physical world in which we live.

1-4 The Scientific Method

The progress of science rests largely on the success of the scientific method in answering the scientist's eternal questions: What? How? Why? When a scientist attacks a problem, he usually examines many phenomena and gradually finds relationships between them enabling him to make a generalized conclusion about them. This reasoning from specific facts to general principles is called *inductive* reasoning. It is the opposite of *deductive* reasoning in which a person states a general rule and then searches for evidence of it in specific facts. Inductive reasoning is the heart of the scientific method.

Much has been said about *the* scientific method. However, there is no set procedure by which scientists go about solving their problems. There are probably as many different methods as there are scientists. In spite of the diversity of techniques with which a problem is approached, the scientist in tackling a new one often follows a pattern or series of steps. In order to gain an understanding of this pattern and of how the scientist uses inductive and deductive reasoning, let us outline in a generalized manner the five steps of the scientific method:

(1) The scientist states the problem to be solved as precisely as possible.

(2) He searches the written record to find all known information related to the problem. In this step he utilizes the experiences of all who have wrestled with the problem in the past. He may even find that the problem has already been solved.

(3) He conducts experiments or collects information applicable to the problem. This step is perhaps the most glamorous and the one that usually comes to the layman's mind when he thinks of science. It is the essence of descriptive science (see p. 2).

(4) He correlates the data and tries to formulate an answer to the problem. The scientist now uses inductive reasoning in order to go from the particular to the general—to go from many single facts to a general conclusion. If, in

this process, he discovers a principle of behavior that holds true for all situations in which this problem is studied, then he can summarize this principle in what is called a *law*. A scientific law is a statement of what is believed to be the true essence of a natural phenomenon, and when the true essence is discovered, it serves as a guide in the search for more information about the problem. The scientist then uses deductive reasoning—reasoning from the general law to particular cases—to test the validity of the law. If the law fits new facts and successfully predicts what may be found in nature, it becomes well established.

As an example of these four steps, suppose that a scientist wishes to study the problem: Which are the better mousers, black cats or white cats? Once the problem is stated (step 1), the scientist goes to the library and studies all the information he can find on the mousing abilities of black and white cats (step 2). If he does not find enough information to solve his problem, he sets up experiments (step 3) in which he observes black and white cats as they catch mice. In order to obtain dependable information, he will probably set up conditions so that there are no outside influences on the performances of the cats. Then (step 4) he correlates and analyzes all the data he has found. If he finds that black cats are better mousers than white cats and if there is no evidence to indicate that this is not a true behavior of the feline family, then he can formulate the law: Black cats are better mousers than white cats.

The role of science does not end with the finding of a law. In the case of the black cat law above, science goes on to ask the question: *Why* are black cats better mousers than white cats? Now the scientist seeks an explanation for the law, and if he finds one that is generally accepted by other scientists who are studying the same or related problems, he is said to have established a *hypothesis* or a *theory* for the law. It is in the development and testing of the hypotheses and the theories of science that the most important discoveries are made, and those scientists who formulate dependable theories are held in highest respect by their fellow scientists.

Few scientists are primarily engaged in theoretical science. Most of them work within the limits of descriptive science where the results of a typical experiment—the determination of the melting point of a substance, for example—will very likely fit into an already existing theory. The majority of scientific research, therefore, consists of collecting data and correlating them into a theory that has been established previously. In no event is the solution of the problem to be considered final. Some of the most widely accepted theories have been refuted by later, more accurate discoveries. Science continually invites a re-examination of its theories. Since we are a long way from comprehending nature completely, the true scientist keeps his mind receptive to new ideas that will explain natural phenomena more

adequately. He regards theories, not as immutable answers, but as man-made and therefore fallible tools. Let us go on to the final step in the scientific method.

(5) The scientist records and publishes his findings so that others may learn the results. This part of the scientific method accounts for much of the success of science and makes possible step two above. History is replete with the mistakes of scientists and nonscientists who could have avoided errors had there been available an account of earlier findings.

Significant discoveries in science are often achieved with the aid of two rather unscientific agents—chance and intuition. Luck, or the "happy accident," has revealed the way to many breakthroughs in scientific barriers, although you must remember that "chance always favors the prepared mind." The history of science contains many examples of chance discoveries. Kekule's solution to a problem concerning the structure of molecules—a problem that had stumped chemists for a generation—came to him as an idle daydream during a joyride in the park. Becquerel accidentally discovered the radioactivity of uranium when he placed a piece of photographic film near some crystals of uranium salt. Some scientists are fortunate enough to experience a "sudden flash of insight" that helps them solve a problem or carries them over the next hurdle. Those scientists who can harness their intuitive ability and treat scientific problems in this manner are held in high esteem by their colleagues. Most scientists, however, have limited intuitive capacities; they must follow the scientific method step by step, with many false starts and missteps before a solution is reached. Many scientists consider the scientific method to be an extension and refinement of common sense. P. W. Bridgman of Harvard University has defined it as "nothing more than doing one's damndest with one's mind, no holds barred."

1-5 Physical Science as a Profession

In America approximately one person in 1500 is a physical scientist. Six per cent of these scientists are engaged in teaching and training new scientists. Over 70 per cent are employed by manufacturing, mining, and construction industries, where they are concerned with research and development, testing and analysis, administration and production. The rest are employed by the government in its various services to agriculture, business, and industry. Although the number of scientists has skyrocketed during the past generation, the demand still far exceeds the supply mainly because industry, government, and academic institutions have found scientists increasingly necessary in their operations. In our scientific society, few technical fields can advance without the constant assistance of physical scientists.

Even though this is an age of specialization, all physical

scientists must obtain a sound training, not only in their own special area, but in related ones as well. Thus, a geologist must study physics, chemistry, and mathematics, and an astronomer or a chemist must have a working knowledge of physics and mathematics. In order to become a specialist in one of the physical sciences, a potential scientist should have a minimum of four years of college training in the science of his choice. And often, because of the growing complexity of his subject, he should spend additional time in graduate study. In some fields it is advisable to obtain a doctoral degree or its equivalent before setting out as a specialist.

1-6 Summary

Physical science is both knowledge of the physical world and the process by which this knowledge is obtained—the scientific method.

In general, the five steps in the scientific method are (1) statement of the problem, (2) search of the literature, (3) experimentation, (4) formulation and explanation of the results obtained from the third step, and (5) publication. Much scientific work, however, is performed without formally following all these steps.

The branches of physical science are physics, chemistry, geology, and astronomy, with mathematics serving as an indispensable tool for them all. Many of the advances being made today involve the areas between these branches, in fields such as physical geology and astrochemistry. The success of studies in these areas is striking evidence of the unity of physical science, which has historically been divided into its several branches.

The physical sciences constitute an active area for professional activity, and, in view of the demands of modern society for his services, the physical scientist occupies a vital place in the world today.

EXERCISES

A. KEY TERMS TO REMEMBER

alchemy
astronomy
chemistry
deductive reasoning
descriptive science
geology
inductive reasoning
mathematics
physics
scientific hypothesis
scientific law
scientific method
scientific theory
theoretical science

B. QUESTIONS ABOUT THE NATURE OF PHYSICAL SCIENCE

1. Outline the primary steps usually involved in a scientific approach to a problem.

2. What are the major branches of physical science?

3. Why are publication and communication of scientific results so important to the advance of science?

4. Select a problem from your common, everyday experience and analyze how you solved it. For example, how did you go about buying a pair of shoes, choosing the route to travel from home to work, planning a vacation, accepting or rejecting an invitation to a party, organizing a new business venture, seeking a job. Now compare the steps you used in solving this problem with those of a scientific approach. Did you solve your problem scientifically?

5. Read several current newspaper or newsmagazine articles concerning science and decide which branch of physical science is discussed in each one of them.

6. Interview a physical scientist and write a report listing the major field of his work, and the scientific problems he is working on.

The following list of references is a guide to further reading in the physical sciences. Many of these references will serve as additional sources for the topics that are introduced later in the text.

A. B. Arons and A. M. Bork, *Science and Ideas* (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1964)

J. B. Conant, *Science and Common Sense* (New Haven: Yale University Press, 1951)

K. S. Davis and J. A. Day, *Water, The Mirror of Science* (Garden City, N. Y.: Doubleday & Company, Inc., 1961)

P. Frank, *Modern Science and Its Philosophy* (Cambridge: Harvard University, 1949)

P. Frank, *Philosophy of Science: The Link between Science and Philosophy* (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1957)

G. Holton, *Concepts and Theories in Physical Science* (Cambridge: Addison-Wesley Press, 1952)

D. P. LeGalley and A. Rosen, *Space Physics* (New York: John Wiley and Sons, Inc., 1964)

C. R. Longwell and R. F. Flint, *Introduction to Physical Geology* (New York: John Wiley and Sons, Inc., 1962)

H. Shapley et al., *Readings in the Physical Sciences* (New York: Appleton-Century-Crofts, 1948)

J. F. White, *Study of the Earth* (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1962)

G. Shortley and D. Williams, *Elements of Physics*, 2nd ed. (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1955)

H. E. White, *Modern College Physics*, 4th ed. (Princeton, N. J.: D. Van Nostrand Co., 1962)

P. Frey, *College Chemistry*, 2nd ed. (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1958)

L. Pauling, *College Chemistry*, 3rd ed. (San Francisco: W. H. Freeman & Co., 1964)

J. N. Friend, *Man and the Chemical Elements* (New York: Charles Scribner's Sons, 1953)

Leet & Judson, *Physical Geology*, 2nd ed. (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1958)

The Journal of Chemical Education *and* The Scientific Monthly *are excellent sources of additional references for many of the topics introduced in this text.*

Two

8734

Space, Mass, Time, and Energy

The essence of physical science is contained in four concepts: space, mass, time, and energy. They form the primary characteristics and intrinsic nature of all natural processes. To understand an exploding galaxy, the change of the moon's phase, the rise and fall of the tides, the flow of water in a stream, the behavior of the tiniest parts inside an atom, or, indeed, all physical phenomena, we must become acquainted with these four basic ideas.

In physical science our interest is not so much to find exact definitions for these terms as it is to grasp their meaning in terms of magnitude: How big is a star, how much space is there in the solar system, how long ago were the rocks formed, how much "stuff" is there in the sun, what keeps the earth moving around the sun? In answering these questions, which are typical scientific ones, the physical scientist is concerned with the measurement of space, mass, time, and energy, and many of the fundamental laws and basic principles of physical science have been discovered in the process of such measurement.

If scientists are to work together in solving scientific problems—and the scientific method permits and encourages the pooling of efforts—there must be definite standards of measurement so that work done in one part of the world can be understood and used in other parts of the world. Several scientific organizations have been established whose primary purpose is to set up convenient and well-defined standards of measurement. Among these are the International Union of Pure and Applied Chemistry and the International Union of Pure and Applied Physics, and in each of the major countries there are governmental agencies, such as the Bureau

of Standards in the United States, that coordinate the international definitions with local practice. In the United States, for example, the measurement of length in inches, feet, and yards is defined in terms of the internationally defined unit of length, the meter. Units of the four basic concepts—space, mass, time, and energy—are defined in terms of actual measurements, and the meter, the gram, the second, and the erg each means the same thing in Cario, Egypt as it does in Cairo, Illinois.

2-1 Numerical Abbreviations

Because the magnitudes of many of the phenomena encountered in physical science are very large or very small compared to our ordinary experiences, scientists have devised a mathematical shorthand for expressing these large and small numbers in terms of powers of ten. Thus, one thousand is 10^3, or ten times ten times ten, or ten multiplied by itself three times. One million is ten to the sixth power, or 10^6. The 3 and the 6 in these examples are called *exponents*. Numbers smaller than 1 are designated with a negative exponent: 0.01 is 10 to the negative 2 power, or 10^{-2}. One-millionth is 10^{-6}, or 1 divided by ten 6 times. Table 2-1 lists numerical equivalents of the powers of ten from 10^3 to 10^{-4}.

These numbers may also be read in terms of a shift of the decimal point. When the exponent is positive, the decimal point is shifted that many places to the right (counting from the 1, not from the 10), but when the exponent is negative the decimal point is shifted the same number of places to the left of the 1. Thus, 10^4 is 1 with enough zeros added to the right so that the decimal point can be moved four places to the right—that is, 10,000.

TABLE 2-1 NUMERICAL EQUIVALENTS OF THE POWERS OF TEN

$1000 = 10^3$
$100 = 10^2$
$10 = 10^1$
$1 = 10^0$
$0.1 = 10^{-1}$
$0.01 = 10^{-2}$
$0.001 = 10^{-3}$
$0.0001 = 10^{-4}$

2-2 Length and Distance

In order to describe the size of an object or the distance between two points, we must have a system of measurement for length. Man, since ancient times, has used arbitrary units

of length such as the hand, the foot, the cubit (the length from the elbow to the end of the middle finger), and the furlong. The magnitude of these units is suggested by their relation to parts of the body or, in the case of the furlong, to a certain number of steps or paces. Since these units were selected simply because they were convenient and not because they were identical with all people, they were not standardized in the various countries of the world and there was much chaos in trade caused by the use of so many different measuring systems. Consequently, the French Academy of Science in 1791 urged that a universal standard be adopted. This suggestion resulted in the creation of the metric system which today has become the official measure in many countries (the United States and Great Britain being prominent exceptions, although the inch is defined in terms of the meter). The metric system is used by scientists all over the world, in measuring the tiniest parts of the atom as well as the farthest reaches of the solar system.

The designers of the metric system originally intended that the meter be defined in terms of the earth itself—it was to be equal to one ten-millionth of the distance from the equator to the pole. Based upon the best surveys of the time, a meter was found to be a bit more than a yard (39.37 inches), and this length was accordingly marked on a platinum-iridium bar that became known as the standard meter. The search for an even more precise standard, required by the complexities of modern nuclear science and space travel, has led to the present definition of the length of a meter: 1,650,763.73 times the wave length of the orange-red line in the spectrum of krypton-86. This last definition is accurate to within 2×10^{-8} meter, several hundred times more accurate than the distance between two marks on a metal bar can be measured. Furthermore, it is a definition based on an unchanging aspect of nature that can be reproduced in any properly equipped laboratory.

The metric system, in contrast to the English system, is based upon decimals. Like our system of numeration, the meter is multiplied or divided into series of tens, so that a tenth of a meter is a decimeter, a hundredth of a meter a centimeter, and so on. (Compare this to our monetary system in which a hundredth of a dollar is a cent.) For

TABLE 2-2 METRIC UNITS OF LENGTH COMPARED TO ENGLISH UNITS

1 meter (m)	= 39.37 inches
2.54 centimeter (cm)	= 1 inch
1 millimeter (mm)	= 0.001 meter
1 micron (μ)	= 0.000001 meter
1 Angstrom unit (A)	= 0.0000000001 meter
1 kilometer (km)	= 1000 meters
	= 0.62 miles

measuring the shortest and the longest distance in the universe, the meter is expanded or divided accordingly. Table 2-2 gives some of the more common measures that are derived from the standard meter.

Since metric units are awkward to apply to distances encountered in astronomical studies, astronomers have devised a unit of distance called the *light-year*, which is the distance light travels in one year, or approximately 9,450,000,000,000,000 meters (six million million miles). Thus, it is 4 light-years to the nearest star and 32 light-years to the North Star.

In our study of physical science we will be concerned with almost unbelievably small distances—one ten-thousandth of an Angstrom unit (A) is the width of the nucleus of an atom —and incredibly large distances— it is one hundred thousand million light-years to the farthest galaxy (Fig. 2-1). In our ordinary daily living, we are hardly aware of anything smaller than a micron (tiny dust particles) nor are we usually seriously concerned with things greater than a few meters, or

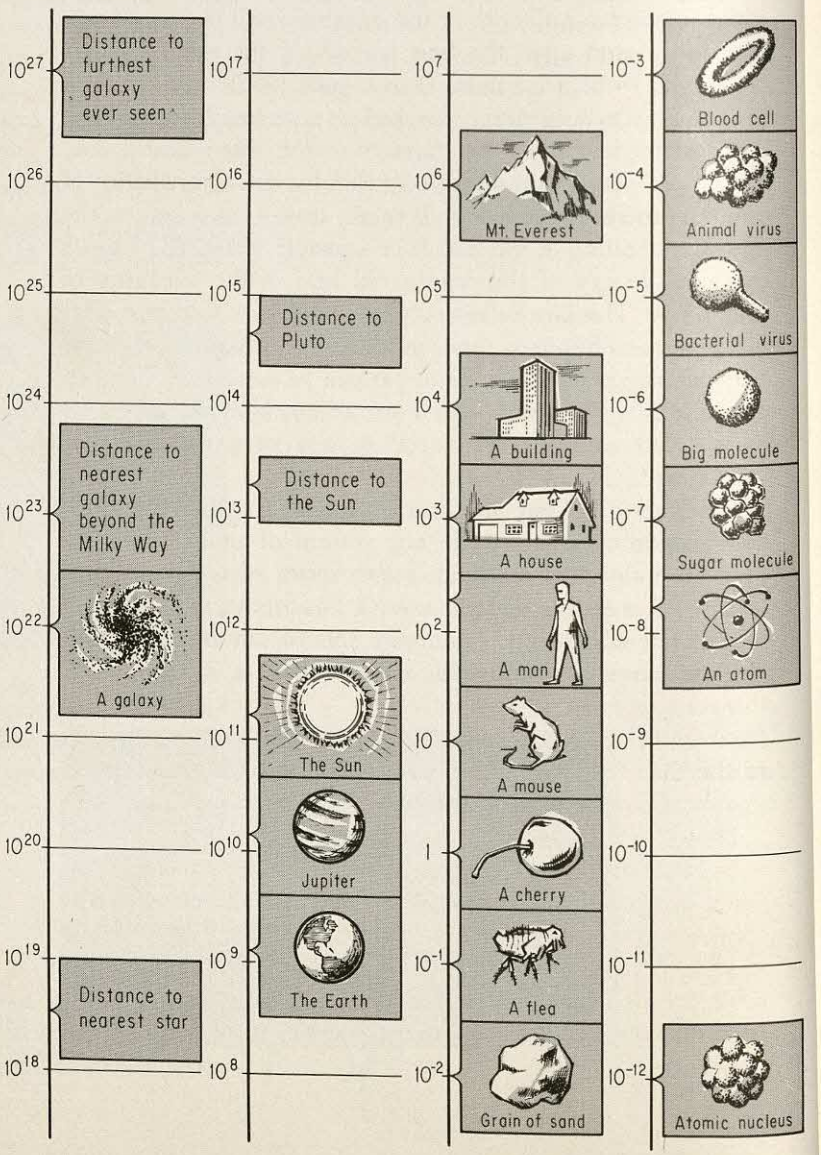


Fig. 2-1. The size of objects from the smallest (the nucleus of the atom—10^{-12} cm) to the largest (the universe—10^{27} cm). (Redrawn by permission from Pauling, *General Chemistry*, 2nd ed. San Francisco: W. H. Freeman & Co., 1953)

at most a few kilometers, in magnitude. In physical science, however, the range of observation and measurement has been extended far beyond the limits of visual perception, and in order to fully appreciate many basic physical principles, we will need to become intimately acquainted with distances that lie outside these limits.

2-3 Area, Volume, and Space

Once we have a universal unit of measure for length, we can then express areas and volumes in universal units. An area is the dimension of a surface or a plane, and in mathematical terms it is the product of its length times its width:

$$a = l \times w$$

The area of a typical playing card is 6 cm times 9 cm or 54 cm^2. We might express the area of a floor in square meters or square feet, the area of a farm in square rods. In all these instances, the area is a measurement that is stated in terms of the square of a unit.

Volume is the size of an object in space. Or, we might say, it is a quantity of space, the mathematical product of three dimensions—length, width, and height:

$$v = l \times w \times h$$

The volume of a sphere is obtained by the relationship:

$$v = \tfrac{4}{3}(3.1416)\, r^3$$

where r is the radius of the sphere. All volume measurements are in terms of the cube of a unit. Thus, the volume of a room may be, say, a thousand cubic meters. Because of the common use of certain volumes, special units such as the gallon (231 cubic inches), the pint, the liter, and so on, have been invented. In the metric system, the standard unit of volume is the *liter*, which is supposedly equal to a cube that is ten centimeters on each edge, or 1000 cubic centimeters. The liter was originally intended to be the volume occupied by exactly 1000 grams of water at 4°C, assuming that each gram of water would occupy one cubic centimeter. But more precise measurements revealed that there was a slight error in this comparison of weight to volume, so that a liter is now defined as the volume occupied by 1000 grams of water in spite of the fact that its volume is actually 1000.028 cubic centimeters. A liter is slightly larger than a quart (see Table 2-4).

These three dimensions in what has been called Euclidean geometry—after the Greek mathematician Euclid of 300 B.C. —not only served to define a volume of space, they are also convenient in describing the location of a point or an object in space. If we state that an object is 300 meters east of a reference point and 400 meters north of the reference point, then we know that it is located some 500 meters in a north-

easterly direction. But, especially in our air age, we may need to know if this point is at the earth's surface or above it. Hence, we need a third dimension to ascertain where this point is an Euclidean space. If an airplane, for example, were flying 150 miles east of Chicago, 200 miles south of Chicago, and 4 miles above Chicago, these three measurements locate the plane for us, but if any of the dimensions were omitted we could not be certain of the plane's position.

In their measurements of space, scientists have found that a man's volume is large compared to the volume of an atom—the diameter of an atom is about 0.00000001 cm—but small compared to the size of the earth or the solar system (Fig. 2-1). Our early concepts of space derived from the observations made with powerful telescopes have, in recent years, been extended by the photographs and other data obtained from spacecraft. The earth is indeed surrounded by a vast space, and it is but a small particle in the solar system. The solar system, in turn, is only a speck in our galaxy, and our galaxy, though its diameter is 100,000 light-years, is only one of many in the universe.

2-4 Mass and Weight

If someone tosses a rubber ball to you, it strikes your hand with a certain force or impact. But if an iron ball of the same size is tossed to you at the same speed, it strikes your hand much harder. The difference in their impact is caused by their difference in *mass*. Even though the rubber ball occupies as much space as the iron ball, the iron ball has more mass because it exerts a greater force on your hand.

Mass can be measured in several ways. On the surface of the earth, we can easily measure the mass of an object by seeing how far it stretches a spring or by balancing it against a known mass on a balance. If one object stretches the spring twice as much as another, we say that the first object has twice the mass of the second one. What we are actually measuring here is the pull of gravity on the objects, and this pull we call the weight. If we took the same objects out into space, however, where there is almost no gravity, their masses would be the same as they are on the earth, but they would have practically no weight. Weight, then, is the pull of the earth's gravity or that of another celestial body on a mass, and this may change, while the mass of the object itself is unaffected by its position in the universe.

Another method of measuring mass involves the impact or force that is associated with the movement of the mass. This method permits us to measure the mass of an object anywhere in the universe. For instance, the difference in the impacts of the rubber ball and the iron ball would be noticeable regardless of where this experiment is performed, because the masses of the rubber ball and the iron ball are not affected by their position in the universe.

2-5 The Analytical Balance

The analytical balance is the most convenient method of accurately measuring mass. It is a lever from the ends of which two pans are suspended. The lever is supported at the middle on a jeweled bearing, and similar bearings support the two pans. The whole system is usually enclosed in a glass case in order to protect the balance from dust, drafts, and temperature changes. Most modern scientific laboratories have analytical balances so sensitive that they can measure the mass of the ink with which you write but one letter of your name. On one pan of the balance a known mass, usually a metal weight, is placed, and an unknown mass is placed on the other pan. Since the earth's attraction for objects of equal mass is the same, a 2-gram mass on one pan will exactly counterbalance a 2-gram mass on the other pan. The analytical balance is an indispensable tool for the modern scientist. Although there have been improvements over the years in the mechanical design of the balance, the primary purpose of this instrument remains the same—the determination of the mass of an object.

2-6 The Metric System of Mass and Volume

The designers of the metric system intended that the only arbitrary unit would be the meter and that all other units could be derived from it. The basic unit of mass, for instance, is the *gram*, which is a quantity of matter equal in mass to one cubic centimeter of water at 4°C where water is most dense. Note the inherent relationship between a linear unit, the centimeter, and the mass unit. If we should lose our set of gram weights, we could readily re-establish them by using the mass of a cubic centimeter of pure water for each gram of weight involved.

Similarly, the unit volume in the metric system is based on the already standardized unit of length. The unit volume is the *liter*, which, as we have seen, is the volume of 1000 grams of water at 4°C. There are a thousand milliliters in each liter, and each milliliter is, for our purposes, identical to a cubic centimeter. In all units of length, mass, and volume there are prefixes that give a multiple or a fraction of the basic unit. This is possible because a decimal relationship is always involved. Table 2-3 gives a summary of these prefixes and the relationships among them.

Both the metric system and the English system of measurement are based on arbitary units, but in the latter we find more cumbersome relationships between units. For instance, 3 feet is a yard, 5½ yards is a rod, and 5280 feet is a mile. In contrast, 10 deciliters is a liter, 100 centimeters is a meter, and 1000 grams is a kilogram. Furthermore, a liter is a cube 10 centimeters on an edge, but a quart is 57.75 cubic inches. Table 2-4 equates some of the metric units with some of our more common English units.

TABLE 2-3 METRIC SYSTEM

Multiple	*Symbol*	*Prefix*
10^{12}	T	terra
10^{9}	G	giga
10^{6}	M	mega
10^{3}	k	kilo
10^{-1}	d	deci
10^{-2}	c	centi
10^{-3}	m	milli
10^{-6}	μ	micro
10^{-9}	n	nano
10^{-12}	p	pico

TABLE 2-4 METRIC AND ENGLISH UNITS

1 meter = 39.37 inches
2.54 centimeters = 1 inch
1 kilogram = 2.20 pounds
1 metric ton = 1000 kg or 2200 pounds
28.35 grams = 1 ounce
1 liter = 1.06 quart

2-7 Density

All of us are acquainted with the relative heaviness of common substances. For instance, we know that lead is heavier than iron, that iron is heavier than aluminum, and that aluminum is heavier than magnesium. For this reason we build airplanes of aluminum and magnesium; our automobiles can be made of iron and steel because other factors, such as cost, rule out the use of the more expensive lighter metals. If we want an object such as a paperweight to be especially heavy, we fill it with a material such as lead.

In scientific work, we need a quantitative means of expressing the heaviness of various substances. This is accomplished by comparing the mass of an object to the volume of space it occupies. This ratio is called *density*. For practical purposes, the most convenient measure of the mass of an object is its weight, particularly for observations and measurements conducted on the earth's surface. Thus the working definition of density is:

$$\text{Density} = \text{Weight divided by Volume}$$

or:

$$d = \frac{w}{v}$$

For example, if 100 pounds of sand occupy 2 cubic feet, then the density of it is 100 pounds divided by 2 cubic feet, or 50 pounds per cubic foot. This equation is valid no matter what material is measured; the density of a boulder is calculated in the same way as the density of air.

8734

In metric units, 50 grams of water occupy 50 cubic centimeters. The density of water, therefore, is 50 g divided by 50 cm^3 or 1 g per cm^3; in the English system, its density is 62.4 lb/cu ft. The density of iron is 7.6 g per cm^3, while the density of mercury is 13.6 g per cm^3. Gold has a density of over 19 g per cm^3. Note that all these density expressions are in terms of weight or mass per unit volume.

The densities of different substances may be compared to each other by assigning to each substance a number that represents how many times heavier or lighter it is than an equal volume of water. This number is the *specific gravity* of the substance. To calculate the specific gravity of say, mercury, we use the following equation:

$$\text{specific gravity of mercury} = \frac{\text{density of mercury}}{\text{density of water}}$$

$$= \frac{13.6\ \text{g/cm}^3}{1\ \text{g/cm}^3} = 13.6$$

Note that when the density of one substance is divided by the density of another, the weight and volume units cancel out and the specific gravity is merely a number without units. If the specific gravity of a substance is less than one, it is less dense than water and so will float in water. The specific gravity of redwood is 0.34 to 0.41, that of certain rocks 2.5 to 4.0, and of cork 0.24.

2-8 Time

In addition to the three dimensions of Euclidean space, we must now add a fourth dimension—*time*. With the general acceptance and successful experimental proof of Einstein's theory of relativity, time is now considered as a dimension that is required along with length, width, and height in describing an event in physical science. With these four dimensions, the position or movement of any object can be accurately described. Even in our ordinary experience, we use the time dimension in describing events. For instance, a football game is scheduled to take place not only at a certain point in Euclidean space but at a certain time.

As human beings, we can alter the location of an event, but there is nothing we can do about the time dimension. If we schedule a football game on Saturday at 2 p.m. we shall have to wait until that time arrives. And once the event has passed, there is no way we can re-create that particular time. We can have a football game at the same Euclidean dimensions, but never at the same time again. This leads us to conclude that time and space are two distinctly separate concepts; attempts to show that they are interchangeable have failed. Even though we may be inclined to kick someone into the middle of next week, it is a physical impossibility.

The true meaning of time is difficult to define. At the least, we can say that it is an interval of duration, but when we ex-

S.C.E.R.T., West Bengal.
Date .....
Acc. No.....

S.C.E.R.T., West Bengal
Library
Calcutta

amine the term "duration" we leave the realm of science. We know that time always moves in one direction—forward. In our minds or dreams we may go backward in time, but in physical reality the succession of events is always forward. Time may seem to pass more quickly on some occasions than on others, but there is no evidence that time moves at any other than a constant rate. In other words, we assume that time moves forward uniformly and relentlessly, and the most we can do is observe and measure it.

There is always an interval between the occurrence of an event and our awareness of it. Light from the moon, for instance, takes over a second to reach us; explosions on the sun occur eight minutes before we actually see them. When we see the sun set in the west, the sun has actually passed below the horizon eight minutes before. If the light emitted from the star Arcturus should suddenly cease, we would not know it until 40 years later.

An interesting correlation between time and space is used in the finding of longitude in navigation. A reference line circling the globe through the two poles, called a *meridian,* is established that runs north and south through Greenwich, England. At twelve o'clock noon, Greenwich civil time, the sun is directly overhead along this meridian. If we move eastward, the sun will be at its zenith earlier than 12 noon Greenwich time (or, moving westward, the sun will be at its zenith later than 12 noon, Greenwich time), and we would be eastward of Greenwich a distance comparable to the earlier time that noon arrives. Since the earth rotates once each 24 hours, we would be east of Greenwich one twenty-fourth of the distance around the earth if noon came one hour earlier than it does at Greenwich. At the equator, our hour is equal to about 1020 miles.

Once the meridian is established, all we need is an accurate clock, set to Greenwich time, that we can carry with us on our travels. At exactly noon by the sun, we note the time and determine our distance east or west of the Greenwich meridian by the time difference between our noon and the noon of the clock. Greenwich time is broadcast by radio and enables us to find our longitude position any place in the world within one quarter of a mile.

2-9 Time from the Sun and Stars

One of the characteristics of time that is of chief interest to the scientist is its measurement. You might think of several possible timekeepers that could be used for this purpose: the ticking of a clock, the beating of a heart, the vibration of atoms in a crystal. An ideal time-measuring device, however, must have two characteristics: it should be unchanging and it should be readily available and convenient to use. Man-made clocks are operated by mechanisms that run down or wear out and are, therefore, not ideal timekeepers. The

heart is untrustworthy because its pulsations vary according to a person's degree of activity and excitement. Since the vibrations of the atoms in a crystal vary with temperature changes and since there are considerable electrical and mechanical problems involved in counting the intervals of their vibrations, the atoms do not make an ideal timekeeper. The nearest thing we have to a perfect timekeeper is the apparent movement of the sun and the stars as the earth rotates on its axis, and this movement is easily observed.

In order to communicate with each other about experiments involving time, scientists have agreed on the second as a reference for accurately describing an interval of time. The *second* is defined as a fraction, 1/31,556,925.9747, of the tropical year 1900. The tropical year is the period of time between successively similar solstices (high point of the noon sun in summer or low point of the noon sun in winter) or the interval between similar equinox (equal length of night and day in either the spring or autumn). The sidereal year, on the other hand, is the period during which the sun returns to the same position in the sky relative to the stars, and this is about 20 minutes shorter than the tropical year. The length of the tropical year 1900 is arbitrarily selected as the reference year because these intervals, including the interval of the daily rotation of the earth, are either becoming longer or are subject to variations which we will examine shortly.

Because it is quite a long time to wait from one year to the next to check our clocks with the tropical year, it is convenient for our professional clock watchers to check the daily rotation of the earth by sighting the stars. The stars do not have to be checked every night; mechanical clocks governed by the vibrations of a quartz crystal can keep time accurately within 0.02 seconds per year. Should the clock run ahead or behind the stars, the necessary correction can be made. These observations are made in this country at the United States Naval Observatory, and these data and those from about twenty other observatories scattered throughout the world are collected at Paris and then broadcast by radio to all parts of the world.

You might ask why we need such accurate measurements of time—in our daily life we usually have at least a few minutes leeway at most occasions. The answer is that the astronomer in his study of the stars often needs an accuracy of one-thousandth of a second. Great precision is also required in the navigation of ships and planes, in the field of communications, and in space science. Furthermore, scientists often pursue the measurement of a physical phenomenon with utmost accuracy merely for the sake of achieving the accuracy. In this search for greater and greater accuracy, scientists occasionally discover a significant relationship that would otherwise have escaped detection.

Accurate measurement of time reveals that the earth is slowing down in its rate of rotation, not to an alarming extent, however, since it amounts to only about 0.001 second per

century. The major reason for this slowdown is believed to be the braking action of the tides on the spinning earth (see Chapter 5). As the earth turns on its axis, the moon attracts the water in the oceans and causes it to lag behind. Consequently, the earth is subject to tidal friction along its ocean beds and shorelines. Other irregularities in the earth's rotation are noticeable, although their causes for the most part are unknown. For example, the earth rotates slightly faster during our autumn than during other seasons of the year, possibly by as much as one-thousandth of a second a day. Perhaps there are certain movements of the atmosphere that account for the speedup, but scientists are not really sure. There is some evidence that the atmosphere undergoes a vertical tidal action caused by the gravitational force of the sun upon the earth's atmosphere. This daily up-and-down effect, coupled with the moon's gravitational force, may have an advancing or retarding effect upon the rotation of the earth's atmosphere, which in turn would affect the rate of rotation of the earth upon its axis.

2-10 Time from the Rocks

The face of the earth presents evidence of great changes that have occurred slowly over extended periods of time. Indications of these long timespans involved in the earth's history can be read in the rocks of the earth's crust. Rocks are known to be formed by sedimentation, a process consisting of the building up of layer upon layer of sand, silt, clay, and other sedimentary materials. By studying these strata, geologists can construct a timetable of the distant past.

It is assumed that the layers of sedimentary rock are formed in a horizontal position and that the top layers were formed after the bottom layers. Earthquakes, upheavals, and deep erosions such as the Grand Canyon in Arizona have exposed layers of rock that to the geologist are as revealing as the rings of a tree. Thus, in Fig. 2-2, the sedimentary rocks at *a* were formed after those at *b*. At *c* the rock strata are inclined, indicating that they were displaced or tilted after they were originally formed and then eroded down to a flat plain. They must have been tilted after they were formed, because sedimentary rocks are always formed in a horizontal position. The events probably occurred in the following order:

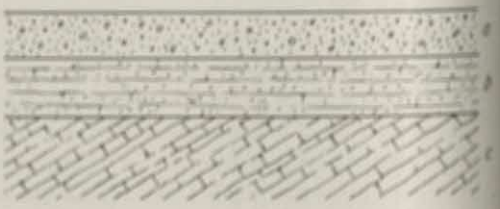


Fig. 2-2. The sequence of events in a rock formation.

1. Formation of the rocks at *c*.
2. Displacement of the rocks at *c* by tilting.
3. Erosion or removal of part of the rocks at *c*.
4. Deposition of the rocks at *b*.
5. Deposition of the rocks at *a*.

By estimating the amount of time that has elapsed between these events, in somewhat the same way as we assume that a

year has elapsed for each tree ring, we can determine approximately when the rocks at *c* were formed.

In recent years it has become possible to measure the ages of rocks with much greater accuracy than by the method described above. This newer method is based upon the analysis of radioactive materials that are found in many rocks. Uranium and thorium are two naturally occurring radioactive elements which change into other elements at a constant rate that can be measured very accurately. We assume that these radioactive processes proceed at unchanging rates because after many observations no measurable change has ever been detected. In fact, scientists have deliberately tried to alter the rate of change by applying heat, pressure, electricity, and powerful chemicals but have never succeeded. Hence we can assume that the same rate of radioactive change has persisted from the time a particular rock layer was formed.

When uranium undergoes radioactive decay, almost all of it becomes lead. The *half-life*, the period of time during which half of the original sample is transformed, can be established for the uranium-lead change, and this period is 7.6×10^8 years. Thus a one gram sample of uranium would, after 7.6×10^8 years be reduced to one-half gram of uranium and the rest of it would have become mostly lead. By analyzing the ratio of uranium to lead in a rock, we can calculate its age. This method, called the radioactive clock, is our most reliable means of measuring geologic time and was used to calculate most of the ages of the periods of time in Table 2-5.

Geologists divide the ancient past into eras, which are

TABLE 2-5 THE GEOLOGICAL TIMETABLE

Era	*Period*	*Dominant life forms*	*Years ago (in millions of years)*
	(Primitive man)		1
Cenozoic	Quaternary	Mammals	17
	Tertiary		58
	Cretaceous	Dinosaurs	100
Mesozoic	Jurassic	"	150
	Triassic	"	214
	Permian	Reptiles	220
	Carboniferous	Amphibians	314
	Devonian	Fishes	355
Paleozoic	Silurian	Sea scorpions	310
	Ordovician	Seaweeds	550
	Cambrian	Trilobites	440
	Late	Soft-bodied invertebrates	800
Precambrian	(Black Hills, U.S.A.)		1800
	(Russian Karelia)	One-celled organisms	1800

separated by major changes in the earth's crust. The Paleozoic era marks the beginning of the formation of large landmasses well above sea level—a time when the processes that laid down the oldest rocks (Precambrian) ceased. The Mesozoic era begins with the upheavals associated with the formation of the Appalachian Mountains in the eastern United States. The mighty upthrusts of the Rocky Mountains in the western United States mark the start of the Cenozoic era. Within each era are subdivisions of time called periods, which are usually associated with the appearance of certain plant or animal forms. For example, in the Devonian period fishes first appeared. Reptiles appeared in the Jurassic period, and mammals made their entrance upon the earth during the Quaternary period. Not only do these periods overlap each other considerably, but many of the plants and animals that first appeared in one period persist in later periods, and many are still living today. The rock record is in agreement with all these periods and eras.

Other scientific analyses also lead us into the past. Sediments such as the red clays found on the floor of the deepest parts of the ocean supposedly increase in depth at the rate of about one millimeter every thousand years. Thus, measuring the thickness of these sediments gives us the approximate age of the ocean floor at that particular place.

The ocean itself gives us a clue to its age through its salt content. About 10^8 tons of sodium in the form of soluble salts are added to the ocean each year by the rivers that flow into it. Since there are 10^{16} tons of sodium now in the ocean —enough to cover the whole earth to a depth of 112 feet— we can estimate the age of the ocean as:

$$\frac{10^{16} \text{ tons of sodium in the oceans}}{10^8 \text{ tons of sodium added per year}} = 10^8 \text{ years}$$

This figure, however, is probably much too small, since we cannot assume that the rivers have always added salt at the same rate as it is being added now and we are not sure that all the sodium that entered the ocean is still there. As in any scientific analysis, we must be certain that all variables are known; otherwise our result is unreliable.

2-11 Time and Man

Table 2-5 shows that man has existed for about two million years—only a moment in the geological timetable. Lower forms of mammals have existed at least two hundred million years, and the oldest rocks on the face of the earth are at least a thousand times older than man. The earliest man that is generally recognized as a member of the human family is the "Ape Man of Java." Fossils have been found that date this man about the middle of the Pleistocene time—possibly half a million years ago. We undoubtedly had earlier ancestors, but there is a lack of agreement among scientists as to which

branch of our family tree accounts for the lineage from which the human race has emerged. The appearance of man more or less as we know him—in height, size, weight, and so forth—took place, many authorities believe, as recently as thirty thousand years ago. This prototype of modern man is the Cro-Magnon man who dwelt in caves in what is now France.

Since man's entrance upon the earth's stage came during the last period of the most recent era, we are in a sense newcomers on this earth of ours; the rocks, plants, and animals that live on the land and in the sea are relatively oldtimers. As a human race, we have just made our entrance into the history of the earth after a long prelude of billions of years. On the other hand, we occupy somewhat of a middle position in size and mass between the galaxies and the atoms. The question of why man is in the middle of the size range but at the most recent part of the time scale is beyond the present knowledge of science. Science has, however, been able to show that the time and space dimensions of the physical world are enormously larger and perplexingly smaller than what our direct senses tell us. These extended concepts of time and space show that modern man occupies a humble place indeed in the world of galaxies, atoms, Paleozoic sea scorpions, and Mesozoic dinosaurs.

2-12 What is Energy?

Energy is another term that defies precise definition, although we use it frequently in our everyday language: certain breakfast foods are advertised as being rich in "energy," our favorite brand of gasoline is full of "energy," a star athlete pursues his goal "energetically," it takes "energy" to climb a mountain, we must now learn how to harness atomic "energy," and so on.

Several types of energy are of fundamental interest to the scientist. The energy of moving bodies is called *kinetic energy*, while the energy an object possesses because of its position (a rock at the top of a hill, for example, has more energy than it does lying at the bottom of the hill) or its state (such as a stretched rubber band or a stick of dynamite) is *potential energy*. A householder buys a certain amount of *electrical energy* when he pays his monthly electric bill. The energy utilized in the burning of oil in a furnace or of foods in our body is *heat energy*. Light and sound are examples of energy associated with wave motion. We will examine these types of energy in detail in subsequent chapters.

The scientist measures energy by examining the effect it produces—that is, he measures the ability of the energy to do *work*. Work, in science, is not used as loosely as it is in common parlance where we may "work" in supporting a weight, "work" on our books, or "work" to get ahead. In

scientific use, work is defined as force multiplied by the distance through which the force operates, or work = force × distance.

This introduces a new term, *force* (to be discussed at length in a later chapter), which for the present you may think of as a push or a pull that can change the position or motion of an object. Force may be measured or expressed in several kinds of units. The smallest unit of force is the *dyne*, which if applied to a gram mass at rest will move it half a centimeter during the first second. This is about the same amount of force expended when we blink an eye. Since the work performed when we blink an eye is the force times the distance, work units may be expressed as dyne centimeters, or *ergs*. Thus, the work performed in blinking an eye is about one erg.

To return to our four basic concepts, if we launched a rocket ship to the moon it would consist of a certain mass (quantity of matter) moving through space (distance) during a period of time (an interval) and possessing a certain amount of energy (ability to do work). As far as the rocket ship itself is concerned, these four concepts—mass, space, time, and energy—just about give us a total picture of its *physical* existence, and that is why these terms are fundamental to physical science. As you proceed in the study of this science, you will encounter these words almost at every turn. Whether in the grand manner of space travel or in the blink of an eye, an understanding of these terms will give you a firm basis upon which to begin your study of physical science. As soon as you become acquainted with their meaning, you will start to feel at home in this field.

2-13 Summary

An understanding of the measurement of space, mass, time, and energy are essential in studying physical science.

The arbitrary unit of length used in the metric system is the meter. With this unit, distance, area, and volume can be described. The unit mass, the gram, is equivalent to one cubic centimeter of water at 4°C, and the unit volume, the liter, is the volume of a kilogram of water.

Mass is a measure of the quantity of matter in an object, while weight is the force of gravity acting upon the object.

Extremely large and small quantities can be expressed in terms of exponential powers of ten. For very large and very small distances, scientists employ special units such as the light-year and the Angstrom unit.

Density is mass per unit volume. Specific gravity is the ratio of the density of a substance compared to the density of a standard, usually water.

Time is an interval of duration, the standard unit for which is the solar second. The second is most accurately determined by observing the apparent motion of the sun and stars.

The geological timetable is divided into periods and eras. The periods represent times of major mountain formations, and the eras mark the appearance of certain animal and plant forms. The most accurate geological timekeeper that we have is the radioactive clock.

Energy is the capacity for performing work, and work is force times distance.

EXERCISES

A. KEY TERMS TO REMEMBER

area
density
energy
era (in geological time)
exponent
force
gram
light-year
liter
longitude
mass
meter
period (in geological time)
sidereal year
specific gravity
time
tropical year
unit
volume
weight
work

B. QUESTIONS ABOUT SPACE, MASS, TIME, AND ENERGY

1. On the basis of your present knowledge write definitions of space, mass, time, and energy. After you have studied the next few chapters, examine your definitions and analyze them for accuracy.

2. If a particular sediment of red clay on the ocean floor is 200 meters thick, in what geological era did the deposit probably begin to form?

3. Of the methods used to determine the age of the earth's rocks, which one is probably the most accurate?

4. Which of the following are examples of work; which are force?

a. holding a book
b. throwing a ball
c. blinking an eye
d. a falling stone
e. pressing a book against a desk
f. coasting downhill

5. Which of the two, the solar day or the sidereal day, can be measured more precisely? Explain your answer.

6. Taking either side of the issue, draw up a list of arguments on the proposition: The United States Congress should enact legislation that requires the use of the metric system of weights and measures in all business and commercial transactions, both domestic and foreign.

7. Even though an analytical balance would not work out in space beyond the earth's gravity, explain why the balance measures mass and not weight.

C. PROBLEMS ABOUT SPACE, MASS, TIME, AND ENERGY

1. The wave length of blue light is 4800 A. Express this in centimeters.

2. Assuming that our galaxy is a sphere, calculate its volume in cubic light-years. Ans. 5.2×10^{14} cubic light-years.

3. It is presumed that there is at least one molecule of hydrogen in every cubic centimeter of space in the universe. Disregarding the planets, moons, and the sun, find how many molecules of hydrogen there are in a sphere of space that would hold the solar system. The radius of the solar system is 6 billion kilometers.

4. Give the sum of the following quantities in grams: 12 g, 1.025 kg, 350 mg, and 6000 micrograms. Ans. 1,037.356 g.

5. What is the sum of the following numbers: 6×10^{2}, 3.5×10^{-3}, 625, 6×10^{0}?

6. Calculate the density of a block of wood, weighing 30 kilograms, that is 20 cm wide, 10 cm thick, and 2 meters long. What is its specific gravity?

7. If you wished to build a raft of redwood (sp. gr. = 0.35), what volume of wood is required to support a load of 100 kg in addition to the weight of the wood itself? Use only enough wood barely to float the raft and to keep the load above the water.

8. If you are aboard a ship at the equator and noon occurs at 10:00 p.m. Greenwich civil time, where would you be? What would your position be if noon occurred at 6:00 a.m. GCT?

9. Columbus, though navigating mostly by dead reckoning, was able to estimate his latitude position, but lacked the means of estimating longitude position. What type of invention was needed to make possible the estimation of longitude?

10. How many cubic picometers are in one cubic terrameter?

11. If a rock of density 3.0 g/cm^3 is tossed overboard from a rowboat, would the level of the lake rise, fall, or remain unchanged?

12. Try to find a rock formation, such as in a road cut, riverbank, or excavation, that shows evidence of sequential sedimentation. What kind of additional information would you need in order to assign an actual time or period of deposition to a single rock formation?

Three

Gravity

Why does a swinging pendulum mark off equal time intervals? Why does a ball thrown upward gradually slow down, stop, and fall back to earth? What holds the earth together, and what keeps the earth in its orbit around the sun year after year? The answers to these questions, which were sought for centuries from the time of the Greek philosophers until the seventeenth century when Newton found the solution to them, stem from one simple fact: between two masses there is a force that tends to pull them together. This force of attraction, called *gravity,* was first identified and described by Galileo Galilei, an Italian who lived from 1564 to 1642. Others had explored the idea of gravity, but it was chiefly the brilliant thinking and experimenting of Galileo that laid the foundation for the modern science of mechanics.

Although the exact nature of gravity has never been explained, we know many of its characteristics. Gravity is a force and, as we have already seen, a force is capable of causing an object to move. The resulting movement of an object involves the concepts of speed and velocity, and when the speed or velocity of an object is changed, we say the object is accelerated. Hence, gravity may affect the speed, velocity, and acceleration of an object. It is gravity that keeps a pendulum swinging to and fro, and it is gravity that accounts for the shape as well as the compactness of the earth. Gravity holds the earth in its orbit about the sun, and gravity is the force that arrests the upward movement of a ball when it is thrown into the air.

Before we consider gravity, we need to become acquainted with the simpler terms— speed, velocity, and acceleration.

3-1 Speed and Velocity

Speed and velocity are often used synonymously, but, for our purpose, we must distinguish between them. Speed simply describes how fast an object is moving, while velocity describes both how fast and in what direction it is moving. For example, we say correctly that a car has a speed of 50 miles per hour, but to speak of its velocity we must say that it is going east (or some direction) at 50 mi/hr. If the car is going in a straight line, the speed and the magnitude of the velocity are the same. If, however, the auto were to go around in a circle at an even 50 mi/hr, its speed would be constant, but because its direction is always changing, its velocity would also be continually changing.

To measure the speed of a car, we must determine its location at successive intervals. Suppose we place observers at one-mile intervals along a highway to note the time when the car passes each checkpoint and then collect the information from each observer and record it in a table.

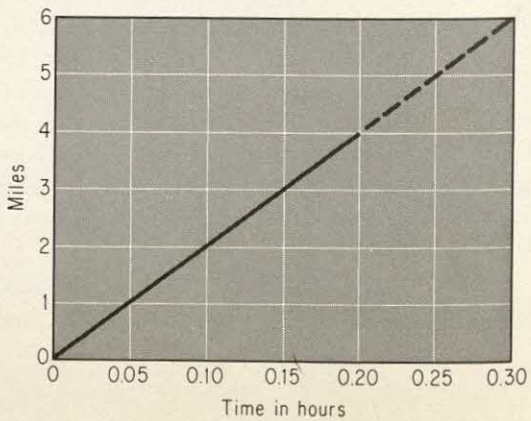
TABLE 3-1

Observer	*Distance from reference point (mi)*	*Time elapsed after car left reference point (min)*	*(hr)*
A	0	0	0
B	1	3	0.05
C	2	6	0.10
D	3	9	0.15
E	4	12	0.20

The table reveals that the car traveled at a uniform speed, since it went the same distance (1 mile) during each 3-minute period. Before we can use this information effectively, however, we must put it into mathematical language, in the form of either a graph or an equation.

In a graph, such as Fig. 3-1, we can plot the two variables involved, which here are distance and time. To do this, we first make some assumptions; in this case, that the car's mass, the model, its driver, and all other factors present in this experiment except time and distance do not affect the velocity of the car. Next we may make the graph, Fig. 3-1, by plotting the two variables of time and distance.

Fig. 3-1. A graphical summary of the data in Table 3-1.



To find a suitable mathematical equation to describe the car's motion, we must find the mathematical relationship that exists between the two variables. Probably the simplest relationship in mathematics is "direct proportion." Are the distance and the time directly proportional to each other? In other words, does the distance increase at a constant rate as the time increases? If so, we can say that

$$s \propto t$$

where s represents distance, t represents time, and the symbol

$\propto$ means "is proportional to." On a graph this relationship, as we saw in Fig. 3-1, is plotted as a straight line sloping upward to the right.

In order to use the actual numbers observed in our experiment, we must express the relationship as an equation. To change $s \propto t$ to an equation, we insert a fixed number, or a constant, that will equalize the two sides of the equation. If we let c be this "constant of proportionality," then

$$s = c \times t$$

Substituting the data from our experiment in this equation, we have:

$$1 \text{ mi} = c \times 0.05 \text{ hr}$$
$$2 \text{ mi} = c \times 0.10 \text{ hr}$$

and so on down the table. Solving one of these equations for the value of c, we get:

$$c = \frac{2 \text{ mi}}{0.10 \text{ hr}} = \frac{20 \text{ mi}}{1.0 \text{ hr}} = 20 \frac{\text{mi}}{\text{hr}}$$

If we solve the other equations, we find that c has the same value in all of them. Substituting this value for c in the general equation gives:

$$s = 20 \frac{\text{mi}}{\text{hr}} \times t$$

Since c is a distance divided by a unit of time, it is defined as a quantity of speed or velocity. Our equation can now be better written as:

$$s = vt \quad \text{or} \quad v = \frac{s}{t} \tag{3.1}$$

where v represents velocity or speed. The equation shows that, when the velocity is a constant, the distance traveled equals the velocity times the elapsed time. Note that this equation is limited; the velocity must be a constant.

Now that we know the equation which relates distance, velocity, and time, we can calculate any one of these factors if we know the other two. For example, suppose the car had continued at the same uniform rate for 15 minutes (0.25 hr), how far would it go in that time?

$$s = vt = 20 \frac{\text{mi}}{\text{hr}} \times 0.25 \text{ hr}$$

$$s = 5 \text{ miles}$$

Would the graph show the same result?

It might be interesting to see how far a car goes in feet each second while it is traveling at 60 mi/hr. We will perform a series of multiplications by the pure number 1. This will not affect the truth of our equation so that we will always have a true equality. Note that we can write

$5{,}280 \text{ ft} = 1 \text{ mi}$ in the form $\frac{5280 \text{ ft}}{1 \text{ mi}} = 1$ and $\frac{1 \text{ hr}}{60 \text{ min}} = 1$, etc.,

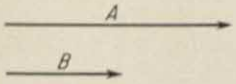
so:

$$\frac{60 \text{ miles}}{\text{hour}} = \frac{60 \text{ mi}}{1 \text{ hr}} \times \frac{1 \text{ hr}}{60 \text{ min}} \times \frac{1 \text{ min}}{60 \text{ sec}} \times \frac{5280 \text{ ft}}{1 \text{ mi}} = \frac{88 \text{ ft}}{1 \text{ sec}}$$

In other words, at 60 mi/hr, you are traveling 88 feet each second. Whenever you must convert from one set of units to another, it can be done in this same way.

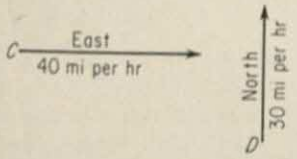
3-2 Scalars and Vectors

In addition to graphs and equations, scientists also use vectors and scalars to record their data. A *vector* is a quantity having both magnitude and direction and may be represented by an arrow. As an example, vector A is twice as great as vector B, although they have the same direction:

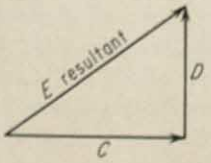


It is convenient to use vectors when we deal with quantities that cannot be adequately expressed by numbers alone. In contrast to a vector, a *scalar* is a quantity that can be completely described by just a number.

Speed, for instance, is a scalar quantity. Velocity, on the other hand, must be a vector since it has both direction and magnitude. Thus, we use vectors in working with velocities. To add scalars is a simple matter: 3 gallons plus 4 gallons equals 7 gallons. But adding vectors requires a special treatment. In order to find the resulting velocity of an object moving at the rate of 40 mi/hr east and 30 mi/hr north, we add vectors C and D, with C representing the first velocity and D the second.



Vectors are added by placing them head to tail successively and then drawing a new vector from the tail of the first one to the head of the last one.



Thus, the resultant velocity, E, is 50 mi/hr in a northeasterly direction. The magnitude of E is determined by measuring its length in the same units as were used to construct vectors

C and D. Or, if the angles between the given vectors are known, the value of E can be calculated geometrically. Here, of course, E is the hypotenuse of the right triangle formed by C and D. Vectors, as we shall soon see, are also used in working with acceleration and force.

3-3 Acceleration

Acceleration is a scientific concept that has crept into our everyday language. An automobile has an accelerator, a pedal that causes more fuel to flow into the engine and thereby produces an increase in the speed of the car. (In the same sense, the brakes might be called decelerators.) In our modern age, we experience acceleration often in our daily travel. When an automobile starts out, especially if it is being driven by a young man with an urgent mission, there is a noticeable thrust and we feel the back of the seat push us forward: our rate of change of velocity is acceleration. Again, as we sit in the seat of a plane about to take off we may feel a tremendous force from the seat on our backs as the plane is accelerated from a standing position to a speed of hundreds of miles per hour in only a fraction of a minute.

Acceleration may be observed in other places. The water that tumbles over the brink of a high waterfall seems to start out slowly, then gain more and more speed and finally descend so rapidly that it perhaps bursts into mist. A stone dropped from a cliff or high building starts out slowly and gradually gains more speed as it falls. A baseball, dropped from the top of the Washington Monument, is accelerated to so great a speed that only the most daring baseball catcher will hazard to catch it. In these examples, it is the force of gravity that causes the acceleration of the various objects.

Acceleration is the rate of change of the velocity of a body as it travels during a unit of time. When a car picks up speed it has positive acceleration; when it slows down it has negative acceleration, or deceleration. And even when the direction of a moving object is changed, acceleration occurs. The reason for this is obvious when we recall that velocity is the rate of motion in a particular direction, so that when we change the element of direction in the velocity we also change the value of the velocity. We experience this acceleration when we round a curve in a car traveling at a constant speed. There is a definite force that pushes us sideways in the seat, a force that may easily be as large as the force produced by accelerating or braking the car on a straight road.

If we let v_0 be the original velocity of a moving object and v the final velocity, then $v - v_0$ is the change in its velocity. Since this change occurs during a time interval, t, the mathematical equation for acceleration is:

$$a = \frac{v - v_0}{t} \qquad (3.2)$$

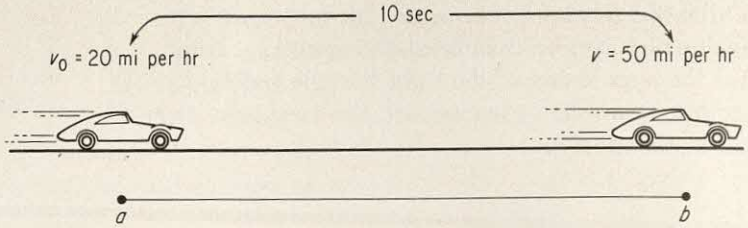


Fig. 3-2. The acceleration of a car.

To illustrate this concept, suppose a car speeds up between points a and b in Fig. 3-2. At a the velocity, v_0, is 20 mi/hr. After a uniform acceleration for 10 seconds, the car is traveling 50 mi/hr. Note the vectors for v_0 and v. The car underwent a change in velocity of 30 mi/hr during 10 seconds. Using Equation 3.2, we find that its acceleration was:

$$a = \frac{50 \text{ mi/hr} - 20 \text{ mi/hr}}{10 \text{ sec}} = 3 \text{ mi/hr/sec}$$

As we saw in the discussion of speed, when any two of three variables are known, the third can be calculated. The same holds true here. If we know the original velocity and the rate of acceleration, we can, by rearranging Equation 3.2, find the final velocity. Multiply both sides of the equation by t and cancel the two t's on the right side:

$$at = \frac{(v - v_0)t}{t}$$

$$at = v - v_0$$

Transposing v_0 to the other side of the equation, we get:

$$v = v_0 + at \qquad (3.3)$$

Thus, if an object with an initial velocity of 100 ft/sec is accelerated 10 ft/sec^2 (10 ft/sec/sec) for 5 seconds, its final velocity is:

$$v = 100 \frac{\text{ft}}{\text{sec}} + 10 \frac{\text{ft}}{\text{sec}^2} \times 5 \text{ sec} = 100 \frac{\text{ft}}{\text{sec}} + 50 \frac{\text{ft}}{\text{sec}} = 150 \frac{\text{ft}}{\text{sec}}$$

The product $a \times t$ can be added to v_0 because they are in terms of the same units—feet divided by seconds.

Both Equations 3.2 and 3.3 give us exact relationships between acceleration, velocity, and time. Equation 3.2 tells us that the acceleration is directly proportional to the change in velocity and inversely proportional to the time during which the change in velocity occurs. Let us consider the acceleration of a racing car that is traveling 100 mi/hr. Suppose that during a minute's time the racer increases his velocity to 180 mi/hr. The change in velocity is 180 — 100 or 80 mi/hr and, since the change is accomplished in one minute, he has accelerated 80 mi/hr/min. In contrast to the racer, a passenger train requires a much greater time to change its velocity, so its acceleration is not nearly as great. Certain animals and birds are capable of changing their velocity quickly and thus have a great capacity to accelerate rapidly. When a golf ball

is struck at the tee, the change in velocity is great and occurs in a minute portion of time. This is indeed rapid acceleration.

Equation 3.3 may be applied to several common experiences. This equation tells us that the final velocity of an object is determined by two factors—its original velocity and the change caused by an acceleration. Unless an object is accelerated, it continues to move at a constant velocity, and if it is continually accelerated, its velocity continues to change. A rocket shot into the air continues to gain velocity as long as there is an accelerating force, but once its fuel is used up, the rocket's velocity change is due to gravitational forces and air friction. Actually, because of these opposing forces, the rocket will be gradually slowed down, or decelerated. It may eventually go to such a great distance from the earth that the gravitational attraction from the sun (or Venus) is greater than that from the earth and its velocity will again be increased. If we knew the values of the acceleration and the time for the rocket's flight, we could calculate its velocity at any time during its flight.

Going back to Equation 3.1, note that it is valid even when the moving object has been accelerated. To apply this equation to an accelerated object, however, we must find the *average velocity* during the time of acceleration. In Fig. 3-2, the average velocity, $\bar{v}$, assuming that the acceleration was uniform, is:

$$\bar{v} = \frac{v + v_0}{2} = \frac{50 \text{ mi/hr} + 20 \text{ mi/hr}}{2} = 35 \text{ mi/hr}$$

Substituting $\bar{v}$ for v in Equation 3.1, we get:

$$s = \bar{v} \times t$$

$$s = \frac{v + v_0}{2} t \qquad (3.4)$$

Although Equations 3.2 and 3.4 are sufficient to solve problems of uniformly accelerated motion, it will be convenient to derive another equation by algebraic manipulation. Solving Equation 3.4 for v gives us:

$$v = \frac{2s}{t} - v_0 \qquad (3.5)$$

Equations 3.3 and 3.5 can now be set equal to each other:

$$v = v_0 + at = \frac{2s}{t} - v_0$$

or:

$$2v_0 + at = \frac{2s}{t}$$

Multiplying both sides of the last equation by $t/2$, we have:

$$s = v_0 t + \tfrac{1}{2} a t^2 \qquad (3.6)$$

Although Equation 3.6 may look far removed from reality, it has been derived mathematically from simpler equations that were obtained from direct experiment. Furthermore,

this equation may be verified by new experiments, proving that the mathematics that were used did not lead us to any false conclusions. Even though mathematical equations may at first appear meaningless, they are useful notations or symbols that say precisely what we mean. For instance, this equation says that the distance traveled by an object is the sum of two factors—its velocity times the time, and half its acceleration times the square of the time it was accelerated. Or we could say that the distance traveled by anything being accelerated uniformly is the sum of two distances: (1) the distance that it would go at the initial velocity if it continued at this uniform velocity for the full time interval, and (2) the distance that it went solely because it was being accelerated during all of this time. We could avoid using equations if we were willing to describe all these relationships in words, but this would require much more space, and it would certainly lead to misunderstandings.

Once the meaning of the symbols and notations is understood, there can be no doubt about what we wish to say. If agreements between individuals and between nations could be written in the form of mathematical equations, there would undoubtedly be fewer disagreements and misunderstandings, but, unfortunately, many of the factors in these matters cannot be expressed in such simple, well-defined terms as those that we have encountered above.

Our interest in deriving Equation 3.6 is merely to arrange old information in a way that will help us understand our next experiment.

3-4 The Inclined Plane

Let us test Equation 3.6 by a simple experiment in which we roll a ball down an inclined plane (Fig. 3-3). As the ball accelerates, we measure the distance it travels during each second. In the first two columns of Table 3-2 we list the time and the distance the ball rolls. Since the ball starts from a dead stop, $v_0 = 0$. Using this value in Equation 3.6, we get:

$$s = \tfrac{1}{2}at^2 \qquad (3.7)$$

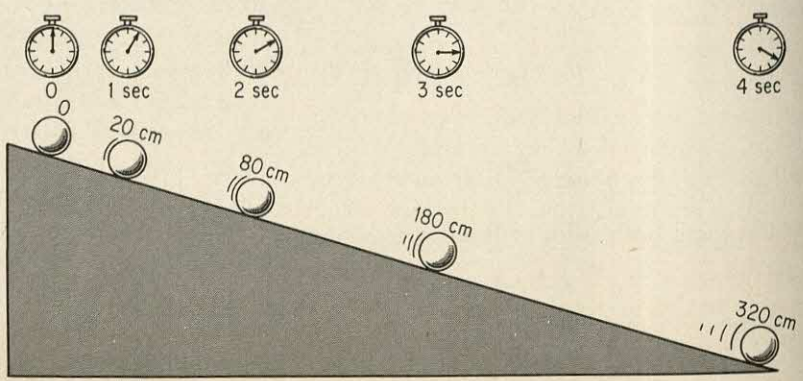


Fig. 3-3. The behavior of a ball rolling down an inclined plane.

TABLE 3-2 DATA OBTAINED FROM THE EXPERIMENT IN FIG. 3-3

Time in sec	*Distance in cm*	*Acceleration in cm/sec²*	*Velocity in cm/sec*
0	0	—	—
1	20	40	40
2	80	40	80
3	180	40	120
4	320	40	160
5	500	40	200

With this equation we can calculate the acceleration in the third column of Table 3-2. Note that the acceleration of the ball is uniform throughout the time it rolls down the slope. In the fourth column, using Equation 3.3, we can record the velocity of the ball.

In Fig. 3-4 let us plot the distance traveled by the ball as a function of the time taken to go this distance. When we join the points for which we have collected data, we have the choice of using a series of straight lines or a single smooth curve. We choose the smooth curve since the ball seemed to move with a uniform acceleration. (You will find that science is full of assumptions, but scientists try to be certain of what each assumption is and to take them carefully into account when they draw conclusions.) We know from the graph that the relationship between distance and time in this experiment is not a direct proportion because its graph is not a straight line. If we draw a line tangent to the curve at the 2-second point, we can find the velocity the ball has at this particular time. In our figure, the tangent shows that the ball at that velocity would have gone a distance of 80 cm (120 cm minus 40 cm) in one second or that it has a velocity of 80 cm/sec.

Since Fig. 3-4 showed that the distance was not directly proportional to the time, suppose we plot the distance versus the square of the time, i.e., 2 seconds will be plotted as 4 (2^2), 3 seconds as 9 (3^2), etc. The result in Fig. 3-5 is a straight line and again we have an indication of a direct proportion: the distance is shown to be directly proportional to the square of the time.

Can we verify the ball's velocity experimentally? If we attach a horizontal plane to the slope at each checkpoint, the ball will roll out along the horizontal plane with the velocity it had at the moment it left the inclined plane. If we measured this velocity, we would find that the results confirm our figures in the fourth column. Actually, in order to make our experiment more accurate we would need to account for the friction between the ball and the plane as well as the wind resistance.

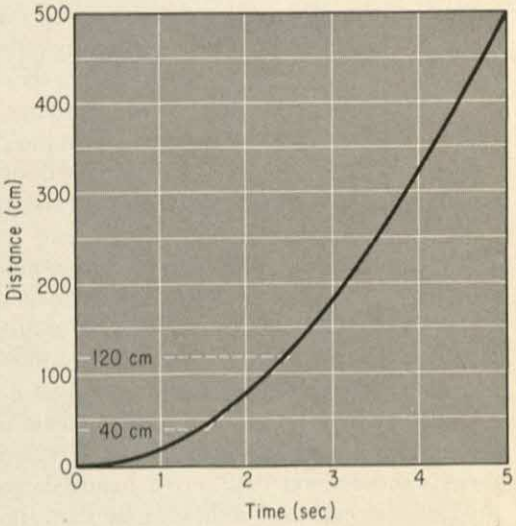
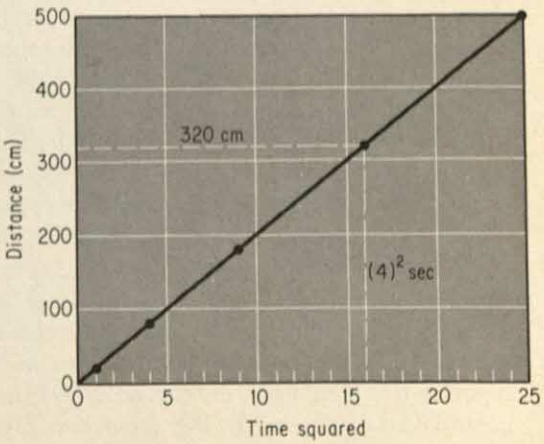


Fig. 3-4. A graph of the data obtained from the experiment of Fig. 3-3.

Fig. 3-5. A plot of the square of the time vs. the distance of the data from the experiment of Fig. 3-3.



3-5 Gravity and Falling Bodies

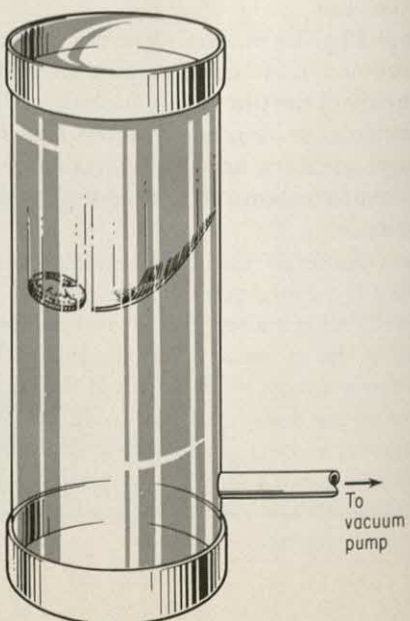
We all know that regardless of where we stand on the earth's surface, our heads will always point away from the earth's center. If the earth were transparent, a man standing at the South Pole would appear upside down to a man at the North Pole; yet both men would be standing in a perfectly normal position. For centuries men have marveled at the mysterious, unexplained force of gravity that holds us all to the earth. We have succeeded in defining many, but by no means all, of the characteristics of this force.

For more than two thousand years men believed that heavy objects fall proportionately faster than light objects. Aristotle, in the fourth century B.C., stated that "the downward movement of a mass of gold or lead, or of any other body endowed with weight, is quicker in proportion to its size." He had apparently observed the effect of air resistance but had not put his full conclusion to any experiment. In the sixteenth century, however, a few men were no longer willing to accept mere reasoning as the truth, and they believed that theories had to be proved by experiment.

In 1590, Galileo, imbued with the inquiring spirit of the Renaissance, challenged the old belief that had been handed down for two millenniums from one unquestioning generation to the next. Although there is no proof that it actually happened, legend has it that he dropped a cannon ball and a small iron pellet from the Leaning Tower of Pisa and found that they reached the ground almost simultaneously. Even though the cannon ball was a hundred times heavier than the pellet, it fell just slightly faster than the pellet. The fact that they fell at so nearly the same rate led Galileo to think something must be radically wrong with Aristotle's theory. After further experimentation, he discarded the theory entirely. His willingness to overlook the slight difference in rates of fall reveals that he had one of the qualities of a great scientific mind—the ability to distinguish between the significant and the insignificant data in experiments.

Shortly after Galileo's experiments, the vacuum pump was invented, enabling scientists to remove air from a container. When a feather and a coin were dropped together in a vacuum, the two objects fell in exactly the same time (Fig. 3-6). From this experiment came the idea that all bodies, in the absence of air, fall at the same rate. Galileo's pellet fell slightly slower than the cannon ball because it had a harder time moving through the air; as Galileo suspected, the earth gives equal acceleration to all objects regardless of weight. An object falls because the force of gravity acts on it, and since at the surface of the earth this force is practically uniform, gravity will always have the same effect upon a 1-pound object. A 2-pound object will have twice the force acting upon it that acts upon a 1-pound object, but these two objects will fall at exactly the same rate because it requires

Fig. 3-6. In a vacuum a feather and a coin fall at the same rate and strike the bottom of the container at the same instant.



exactly twice as much force to accelerate a 2-pound object as it does to accelerate a 1-pound object the same amount.

How does the acceleration of a ball dropped in space compare with that of a ball rolled down a slope? A free-falling steel ball will drop 16 ft in one second (Fig. 3-7), 64 ft in two seconds, 144 ft in three seconds, and so on. By examining the results of this experiment, we find that there is a relationship between the distance and the time involved—they are proportional to each other in the following way:

$$s \propto t^2$$

In order to equate these two quantities, we say:

$$s = kt^2$$

This equation is similar to Equation 3.7 in which $\frac{1}{2}a$ appears in place of k, and the meaning of k is the same as $\frac{1}{2}a$ (both are constant values). In Galileo's experiment with the inclined plane, the accelerating force of gravity was diluted by making the ball acquire a horizontal velocity as it moved downward. The same kind of force operates in free fall and in the inclined plane experiment; the only difference is that in the latter the intensity of the gravitational force is, in effect, reduced by making the ball roll down an inclined plane.

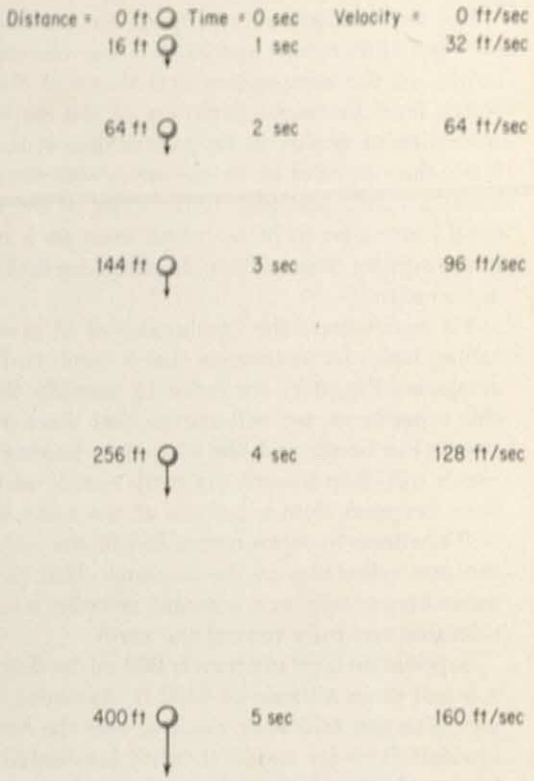


Fig. 3-7. The relationship between distance, time, and velocity for a freely falling object.

3-6 The Acceleration Due to Gravity

Perhaps the most common cause of uniformly accelerated motion is that provided by gravitational force. A rock dropped from a cliff, the trajectory of a bullet shot from a gun, a baseball eventually returning to earth even though knocked out of the ball park, and even our bodies as we walk, all provide examples of a uniformly accelerated motion resulting from the "pull of gravity." This steady pull of gravity depends upon the mass of the object and the gravitational force field of the earth which we will discuss later. Now, before you read on, see if you can determine the acceleration that gravity gave the object in Fig. 3-7.

Neglecting air resistance, which would greatly complicate the problem but which would make little real difference on a compact body such as a rock, we find the acceleration due to gravity to be 32 ft/sec^2 or 9.8 m/sec^2. Substituting some of the data of Fig. 3-7 in Equation 3.7, we get:

$$s = \tfrac{1}{2}at^2$$

$$144\text{ ft} = \tfrac{1}{2}a\,(3\text{ sec})^2$$

$$a = \frac{144\text{ ft} \times 2}{9\text{ sec}^2} = 32\text{ ft/sec}^2$$

Would other values from Fig. 3-7 give the same results for acceleration? They should, if the acceleration is uniform.

The acceleration of gravity on the earth's surface varies between 32.09 ft/sec^2 and 32.26 ft/sec^2 because of the irregularities in the composition and shape of the earth. At the North Pole, the slight flattening of the earth causes the acceleration of gravity to be greater than it is on the equator. Since the weight of an object varies with the gravitational attraction (and consequently the value of the acceleration), it would be easier to lift a pound mass on a mountain top or at the equator than at the pole, assuming that all other factors are constant.

To demonstrate the continual pull of gravity on a freely falling body, let us imagine that a bomb is dropped from an airplane (Fig. 3-8). In order to simplify the conditions of this experiment, we will assume that there is no friction between the bomb and the air. After leaving the plane, the bomb will drop toward the earth exactly as fast as if it had been dropped from a balloon at the same elevation, and it will continue to move forward with the same speed and in the same direction as the airplane. That is, the bomb will move horizontally at a constant velocity, $a = 0$, while it accelerates vertically toward the earth.

Suppose an airplane travels 600 mi/hr due west and drops a bomb at an altitude of 6400 ft. Assuming that there is no air resistance, how long would it take the bomb to strike the ground? How far would it travel horizontally? What would be its velocity as it struck the ground?

By using Equation 3.7, we can find out how long it would take the bomb to hit:

$$s = \tfrac{1}{2}at^2 \quad \text{or} \quad t^2 = \frac{2s}{a}$$

$$t = \sqrt{\frac{2s}{a}} = \sqrt{\frac{2 \times 6400 \text{ ft}}{32 \text{ ft/sec}^2}} = 20 \text{ sec}$$

Thus, it would take 20 seconds for the bomb to fall to the earth.

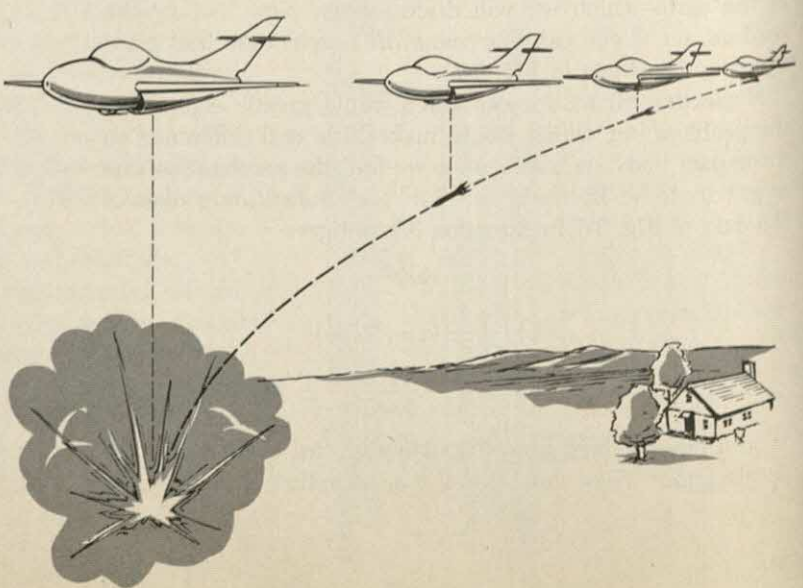


Fig. 3-8. The path of a falling bomb released by an airplane that is traveling with a uniform velocity.

You will notice that the fact that the bomb had a forward velocity made no difference in its time of fall. It is the gravitational force, directed toward the center of the earth, which alone caused the downward motion. Since this accelerating force was downward, it made no change in the forward motion of the bomb.

The bomb's maximum downward velocity, caused by the force of gravity, may be found from Equation 3.3:

$$v = v_0 + at \quad \text{but} \quad v_0 = 0$$

$$v = at$$

$$v = 32 \frac{\text{ft}}{\text{sec}^2} \times 20 \text{ sec} = 640 \frac{\text{ft}}{\text{sec}}$$

The direction of the velocity vector is continually changing during its free fall and is the direction the bomb would be traveling at any point during its fall.

During all the time of fall, 20 sec, the bomb was traveling forward at its original 600 mi/hr. The horizontal distance it covered is given by Equation 3.1:

$$s = v \times t$$

$$s = 600 \frac{\text{mi}}{\text{hr}} \times 20 \text{ sec}$$

But both hours and seconds are time units and we should use either hours or seconds so these units will cancel and we will be left with just the unit of distance. So, using the same process referred to in Section 3-1, we change 20 seconds to hours and,

$$s = 600 \frac{\text{mi}}{\text{hr}} \times 20 \text{ sec} \times \frac{1 \text{ min}}{60 \text{ sec}} \times \frac{1 \text{ hr}}{60 \text{ min}} = 3.36 \text{ mi}$$

Both the airplane and the bomb, therefore, would travel 3.36 miles westward. In other words, the airplane would always be directly above the falling bomb and it would be directly above the explosion when the bomb struck its target. Remember that, in eliminating air resistance from the experiment, we are creating a somewhat unrealistic condition. In an actual situation, the plane would pull ahead of the bomb because friction with the air would slow down both the horizontal motion and the ideal downward velocity of the projectile.

Another example of the continuous effect of gravity upon falling objects is shown in Fig. 3-9. Here a gun, E, is fired at a target, A, which is dropped from its support at the same instant the bullet leaves the gun. The target falls along the line AB, and gravity causes the bullet to fall to the earth along the line CD. The bullet strikes the target where CD crosses AB because the accelerating effect of gravity operates on all objects regardless of their horizontal motion. Would the bullet strike the target if the gun were aimed from a lower position, say at F?

3-7 Mathematical Formulas

In our analysis of various experiments, we have found it necessary to use simple mathematical equations to express our observations. We showed, for instance, that when two quantities, a and b, are *directly proportional* to each other, we can write:

$$a \propto b$$

$$a = k \times b$$

When two quantites, c and d, are *inversely proportional* to each other—that is, c increases as d decreases—we can describe this relationship in this way:

$$c \propto \frac{1}{d} \quad \text{or} \quad c = k \frac{1}{d}$$

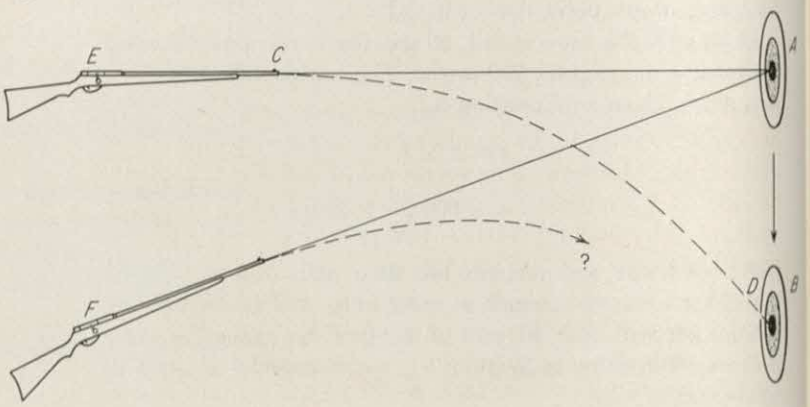


Fig. 3-9. The effect of gravity upon a projectile.

The last equation may be rearranged to give:

$$c \times d = k$$

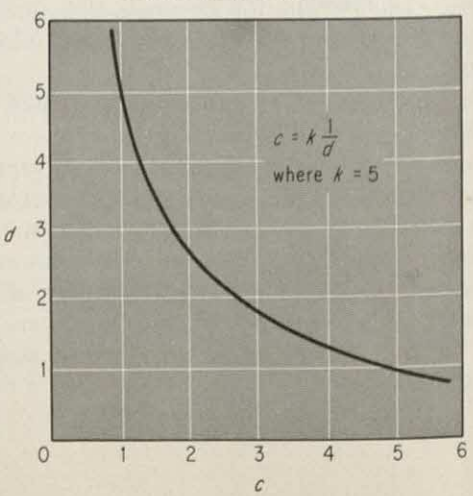
A graphical plot of this inverse proportion produces a curve called a *hyperbola*, which is illustrated in Fig. 3-10.

When a quantity is *proportional to the square* of another quantity and the relationship is plotted on a graph, the resulting line is called a *parabola* (Fig. 3-11). The general equation for a parabola is:

$$e = k \times f^2$$

We have already encountered several mathematical formulas that are illustrations of these three relationships. Equation 3.6, for instance, tells us that the distance that an object travels varies directly with the square of the time during which the object is accelerated. If we plot distance against time, we will have a parabola, the precise shape of which will be determined by the value of the acceleration constant. For many everyday experiences, we can plot relationships of a similar sort. If we drive a car at a constant speed, a plot of the distance against the time would be a straight line sloping upward. If the car is accelerated, its performance might be described by a parabola, such as in Fig. 3-11. Graphic repre-

Fig. 3-10. A graphical representation of an inverse proportion.



sentation is a convenient means of visualizing a mathematical relationship, for it gives us a picture of how two quantities are related to each other. We have often seen graphs used to show the ups and downs of the business cycle but in science they are commonly used to give a concise, visual analysis of an experiment.

With an understanding of these three basic relationships—direct proportion, inverse proportion, and proportional to the square—we can now examine one of the greatest discoveries ever made in science—the discovery of the universal law of gravitation by Sir Isaac Newton in the seventeeth century.

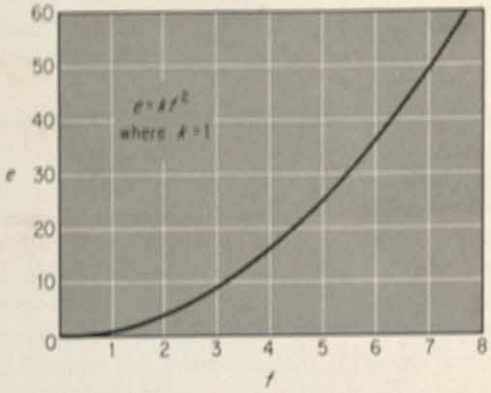


Fig. 3-11. A graph of the equation $e = kf^2$ where $k = 1$.

3-8 Newton's Law of Gravitation

Isaac Newton, born on Christmas Day, 1642, was a quiet farm boy who, like the typical modern boy, enjoyed playing with mechanical toys. He grew up with a strong liking for mathematics and throughout his school and college days excelled as an eager, intelligent student. By the time he was in his early twenties he had made some of the most advanced discoveries of his time, discoveries which still stand as classic theories and principles in the areas of mechanics, heat, and light. He invented his own system of mathematics—the binomial theorem and differential calculus—in order to more adequately describe his ideas, and these new mathematical operations have since become standard tools in the advanced studies of all physical sciences.

The story of how Newton sat under an apple tree and was struck on the head by a falling apple is familiar to us all. Although this incident is probably more legendary than factual, he did, in 1687, express the force of gravity in quantitative terms in his famous treatise, *The Mathematical Principles of Natural Philosophy*, more commonly known as the *Principia*. Undoubtedly the concept of gravitation as a force had been recognized for some time prior to Newton. Galileo, you will recall, published his account of the effect of gravity on falling bodies some forty years before the *Principia* appeared. Newton, however, succeeded in giving a precise explanation of how the force of gravity varies, and his brilliant analysis established him almost immediately as one of the most able scientists in history.

Newton based his study of gravitation upon several important facts that were known at the time, but in addition to knowing these facts, he had an almost uncanny ability to see the fundamental relationships behind them. In relating the moon to the apple, Newton's mind may have followed a series of thoughts such as this: the moon moves in a well-defined path around the earth, and the earth moves in a definite path around the sun. Why doesn't the moon fall to the earth as the apple did? Maybe it does. Some force must hold such large objects in these positions month after month and year after year, and there must be some kind of force

only magnitude, it is a scalar. Weight, on the other hand, is the result of a force from a particular direction and therefore is a vector. Let us see if we can illustrate this difference.

In Fig. 3-13, a mass m is permitted to fall freely and, of course, the force of gravity will accelerate its fall. Since, as we learned from Newton's law of gravitation, the force of gravity, F, is proportional to the mass and since it is also proportional to the acceleration, we can say that:

$$F \propto ma$$

The mathematical reasoning here is very much like saying that the area of a surface is proportional to two factors—the length and the width. We are all familiar with the mathematical equation that says that the area of a surface equals the product of its length times its width. Thus, since force is proportional to both the mass and the acceleration, we can say:

$$F = ma \tag{3.9}$$

This equation is true provided we use the appropriate units. If we express mass in kilograms and acceleration in meters and seconds, then the force is in newtons. Notice that 1 newton is the force which will accelerate a 1-kg mass 1 meter per second per second (1 m/sec^2).

For a 1-kg mass, accelerated in a free fall by the force of gravity where:

$$a = 9.8 \text{ m/sec}^2$$

the force will be:

$$F = 1 \text{ kg} \times 9.8 \frac{\text{m}}{\text{sec}^2} = 9.8 \text{ newtons.}$$

In other words, 9.8 newtons of upward force are required to lift a 1-kg mass at constant velocity or support it at rest; as long as we continue to apply 9.8 newtons of force upward, the 1-kg mass will remain supported and at rest if it was originally at rest. The upward force that we must exert on any object in order to carry it may be calculated in a similar fashion.

Whenever we lift an object, we usually think of its weight. Actually, the weight of an object is a downward force that is very closely related to the force required to lift the object. Equation 3.9 is usually modified slightly when dealing with weight in the following manner. Substituting W for F and substituting g, the acceleration of gravity, for a in Equation 3.9, we get:

$$W = mg \tag{3.10}$$

This equation is a specific application of the general Equation 3.9 which we shall see more of later. Equation 3.10 describes precisely the meaning of the weight of an object. It tells us that weight is a downward force vector (downward because

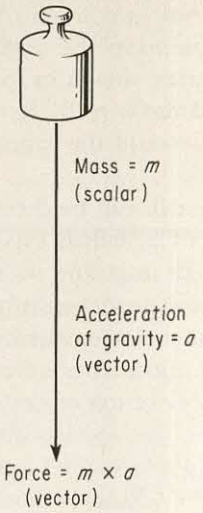


Fig. 3-13. A free-falling object, showing scalar (mass) and vector (acceleration, force, and weight) quantities.

the acceleration of gravity g is in a downward direction) and directly proportional to the mass of the object. Since g is almost constant anywhere on the earth's surface, the weight of an object varies primarily with its mass.

The confusion between mass and weight arises from the use of certain units. If we use metric units, there is no confusion—a 1-kg mass has a weight of 9.8 newtons. However, in common usage, as well as in engineering, the value of g is eliminated in Equation 3.10, and we usually speak of a 1-pound mass as having a 1-pound weight (here we let $g = 1$). It would be more accurate, and probably less confusing, if we were to show the weight of a 1-pound mass in terms of Equation 3.10:

$$W = 1 \text{ lb} \times 32 \frac{\text{ft}}{\text{sec}^2} = 32 \text{ poundals}$$

The *poundal* is the absolute unit of weight in the English system that compares to the newton in the metric system. A 160-pound man, according to this viewpoint, would have a weight of 160 lb $\times$ 32 ft/sec^2, or 5120 poundals. As you can see, there might be a psychological advantage in listing the weights of our football players in poundals.

With this definition of weight, we can now appreciate the interest in some quarters in finding a device or method for reducing the acceleration of gravity. These so-called "antigravity machines," popular in science fiction, would decrease the force of gravitational attraction, and hence the acceleration of gravity, to zero. In practice, this is achieved when we have equal but opposite forces which exactly balance the gravitational attraction. This occurs in an orbiting satellite or in a high-speed plane flying in a special curved path where the forces involved in circular motion (discussed in Chapter 5) provide the proper conditions. The astronaut has mass, with its inertia, but he has no weight. This whole question of weightlessness and its physiological effects is interesting and important in our space program.

Since a mass has no weight unless there is a force acting on it, we find when we examine a particular mass out in space, beyond the practical effects of any gravitational force, that it has no weight and, if it were not subjected to any accelerating force, its motion would not change. When we stand on the face of the moon, we will observe a dramatic example of the effects of gravity. The moon's gravity is only one-sixth that of the earth. A man, therefore, who weighs 150 lb on the earth will weigh only 25 lb on the moon, although his mass, of course, would remain the same. If this man could jump 6 ft high on the earth, he could jump 36 ft high on the moon. You can imagine what would happen to Olympic records if the next games were held on the moon; pole vaulters would be flinging themselves over houses and broad jumpers would be leaping across the width of a football field.

3-10 Summary

Speed is rate of movement, and velocity is rate of movement in a particular direction. A scalar is a quantity that has only magnitude, and a vector is a quantity that has both magnitude and direction. Consequently, speed is a scalar and velocity is a vector.

Acceleration is the rate of change of the velocity of an object; it is the change in velocity per unit of time. The distance that an object moves is governed by the velocity and the time during which it moves. If the object is accelerated, the distance is changed in proportion to the acceleration and the square of the time during which the acceleration occurs.

Gravity is a universal force of attraction between two bodies. The force of gravity is proportional to the masses of the two bodies and inversely proportional to the square of the distance between their centers. The force of gravity accounts for the acceleration of falling bodies.

If the force of gravity between two bodies, their distance apart, and the mass of one of them is known, the mass of the other can be found. This is one method of finding the mass of the earth.

Weight is the effect of the force of gravity upon a mass.

EXERCISES

A. KEY TERMS TO REMEMBER

acceleration	scalar	velocity
gravity	speed	weight
mass	vector	

B. QUESTIONS ABOUT GRAVITY

1. State the universal law of gravitation.

2. Would an analytical balance detect a difference in the acceleration of gravity at two different points on the earth's surface? Would a 1-g mass weigh more, less, or the same at the North Pole and at the equator?

3. How would you expect the gravitational attraction to vary between points at: the top of a mountain and the base of the mountain; differing latitudes; at the earth's surface over a lead mine and over a salt mine?

4. If we drop a rock down a deep vertical mine shaft, it will strike a point very slightly to the east of its starting point. Why?

5. When one of our deep space probes is said to have "passed beyond the earth's gravity," what is meant?

6. If one of our astronauts experiences 8-g acceleration, what do you conceive this to mean?

C. PROBLEMS ABOUT GRAVITY

1. In the example illustrated by Fig. 3-1, what is the numerical value of the "constant of proportionality" for the equation at the

point where the time is 0.15 hr, 0.3 hr? What would be the numerical value of this constant if its units had been mi/min?

2. Draw a graph showing the approximate relationship between two quantities that are directly proportional to each other. Make another graph for two quantities that vary inversely with each other.

3. Use one of the equations in this chapter to: (a) determine how far a car, moving at 50 mi/hr, would go in 3 hours; (b) how far in feet would this car go in 1 second?

4. If a 20-lb horizontal force and a 30-lb vertical force upward act upon an object, which way will it move? Make a vector diagram for this problem and show the magnitude of the resultant force.

5. From your understanding of graphs, what would you predict would be the form of the graph of the data from Table 3-2 if we plotted the velocity of the ball against the time?

6. On Fig. 3-4 show that the velocity at the end of 4 sec is 160 cm/sec.

7. State whether the following are examples of direct or inverse proportions: (a) 8 oz of beans cost 20 cents, 16 oz of beans cost 40 cents; (b) my bank balance and the amount I spend; (c) the pressure of my foot on the car's accelerator and the velocity the car will attain; (d) the sharpness of a curve and the speed a car can go in rounding the curve.

8. Plot the area of a square surface against the length of one side of the square. What kind of a curve is the result?

9. If, for all practical purposes, the earth's radius is 4000 miles, what would you estimate the acceleration due to the earth's gravitational field would be on a satellite at a point 4000 miles above the earth's surface. Ans. 8 ft/sec^2.

10. What would be the acceleration toward the earth, due to the earth's gravity, of an object as far away as our moon (say 240,000 miles away from the earth's center)?

11. An airplane heading east at 120 mi/hr, is blown off course by a 30-mile wind from the south. What is the actual velocity of the plane with respect to the ground?

12. One man can pull with a force of 100 lb, another can pull 150 lb. If they pull together what force can they exert on an object? If they pull in opposite directions what will their force be which might cause the object to move? If they pull on the object but there is 60° difference in the direction they pull, what will be the amount of their combined pull on the object?

13. If a golf ball is in contact with the club of a driver only one-thousandth of a second and achieves a velocity of 100 ft/sec as it leaves the tee, what is its acceleration?

14. If you dropped a stone from a high cliff and 6 seconds later it reached the bottom, how high is the cliff? Ans. 576 ft.

15. If you were to land on a planet that had the same diameter as the earth but a mass ten times greater, what would be your weight?

16. If the diameter of the moon is 2160 miles and its density is 3.3 g/cm^3, what is its mass in tons? Ans. 8.25×10^{19} tons.

17. Using the graph, Fig. 3-14, determine:
    a. What was the car's velocity at the start ($t = 0$)?
    b. What was the acceleration of the car in the first second, in miles per hour per second?
    c. What was the car's acceleration during the third second? the fifth? the sixth?
    d. How fast was the car going in miles per hour after the first two seconds?
    e. How far did the car go during these first 7 seconds?

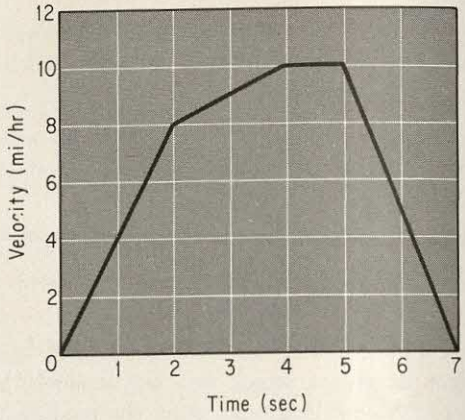


Fig. 3-14.

Four

Gravity in Natural Processes

When we travel through the countryside, we usually consider the landscape—the hills and valleys, the mountains and plains—to be quite solid and permanent. During the course of a lifetime, we ordinarily see little change in the contours of the landscape: rain falls, streams flow, and winds blow, with little apparent effect upon the shape of the earth's solid face. Yet, if we examine carefully the geological evidence that exists on almost every hand, we find that over a long period of time vast changes have occurred in the earth's surface. Deep scars here and heaps of debris there tell us that mighty forces have acted upon the earth's topography. River canyons, eroded gullies, and the long gash in a hillside left by a rockslide are but a few of the indications that the earth's surface is being constantly transformed. What are the forces that cause these changes?

In the last chapter, we examined some of the fundamentals of the force of gravity—its effect upon objects and the mathematical equations that describe its behavior. We saw that gravity affects an object at all times, regardless of its horizontal motion; a bomb dropped from an airplane and a bullet fired from a gun are constantly accelerated toward the earth. Gravity acts in the same way upon rain, snow, and hail, causing the latter to fall at times with destructive violence.

4-1 Gravity as a Force in Erosion

Raindrops usually fall with a velocity of about 30 feet per second, and they would fall faster if it were not for friction with the air. Even so, the pounding action of one-tenth of an

inch of rainfall with this velocity on an acre of ground—11 tons of water in all—is equivalent to almost 100 horsepower. About 90 per cent of this energy is dissipated as the droplets strike the ground, but the remaining portion may cause considerable damage to the earth's surface. The amount of soil that is detached from its normal position depends upon the size of the droplets, the character of the soil, and the amount of cover that protects the soil. On forests and grasslands, for instance, falling raindrops remove very little soil, but on a newly cultivated cornfield, up to 150 tons of soil have been removed from an acre by a 3.8-inch rainstorm. Fortunately, most of this soil was redeposited on lower areas and, though badly damaged, not completely lost.

In addition to the vertical force of raindrops, the rain that collects in streams and rivers exerts a horizontal force that does immeasurably more damage than falling drops. Depending upon the type of soil upon which the rain falls, up to 25 per cent of the rain may drain into streams, the rest being absorbed into the ground. It has been estimated that 30,000 cubic miles of water fall on the land areas of the world each year, and that of this volume 8000 cubic miles run off through streams and rivers into the oceans. The energy of all this flowing water is equivalent to nearly 11,000,000,000,000 horsepower-hours—approximately the energy released by 2000 super hydrogen bombs—and some of it is used to detach soil and rock particles from the earth and to wash them down to the sea. Another way of visualizing this great amount of energy is to consider the work required to accomplish one horsepower, that is, to lift a pound mass 550 feet in one second. A horsepower-hour is one horsepower operating continuously for one hour, or 3600 seconds. Since a horsepower-hour would lift 3600 pounds 550 feet in the air, then 11×10^{12} horsepower-hours would have energy to lift 5.5×10^{9} tons of earth upward 550 feet. What is the source of all this energy? In Chapter 3 we discussed the experiment in which

Fig. 4-1. Two large alluvial fans deposited at the base of the Inyo Mountains in Owens Valley, California.

a ball was rolled down an inclined plane, and we saw that the force that powered the rolling ball was the acceleration of gravity. Since a stream bed is nothing more than an inclined plane, a stream of water flowing down a hillside or through a gently sloping valley is also powered by gravity.

The average height from which the water in the world flows off the land is about half a mile above sea level. If the water dropped straight down to the sea, it would obviously reach a tremendous velocity. Even if the water in a stream were accelerated by as little as 2 per cent of the acceleration of gravity for a period of one hour, its velocity would become over a thousand miles per hour. No stream, of course, moves more than a few miles an hour, because much of its enormous energy is used up in the form of friction with its bed and its banks. It is this friction that causes the erosion of the land into the sea.

What happens to the particles of soil and bits of rock that are picked up by a stream? As long as the stream moves rapidly these particles remain suspended, but when the stream reaches a plain where the incline is less steep, the water slows down and drops the larger particles to the bottom, where sand and gravel bars are often formed. In steep mountain areas, streams often move rapidly until they reach the edge of a broad valley, where they dump part of their load of sediment in wedge-shaped deposits called *alluvial fans* (Fig. 4-1). As a stream empties into the ocean, the still slower flow of the water causes even the finer particles to settle out into a *delta*. There are numerous deltas throughout the world, varying in size from the tiny one formed by little streams to the mammoth one at the mouth of the Mississippi River. Even these large deltas are forming so rapidly that their increasing size can be observed from year to year. If the stream carries only a small amount of sediment and if it is moving rapidly as it enters the ocean, it is not likely to form a delta.

A stream, in short, is forced downhill by gravity and in the process becomes a powerful agent of erosion, picking up deposits when it runs swiftly and dropping them when it travels slowly.

4-2 Erosion and the Landscape

One need travel only a short distance in any country to observe land forms that are produced by erosion. Let us examine the fundamental stages of stream erosion and consider some examples that illustrate these stages. A stream usually evolves through three periods—youth, maturity, and old age. In youth (Fig. 4-2), it cuts a deep V-shaped valley and washes away only the solid material that is within its immediate reach. The scouring action of the stream upon its own bed and banks, plus the collapse of material from the banks into the stream, accounts for most of the erosion. The

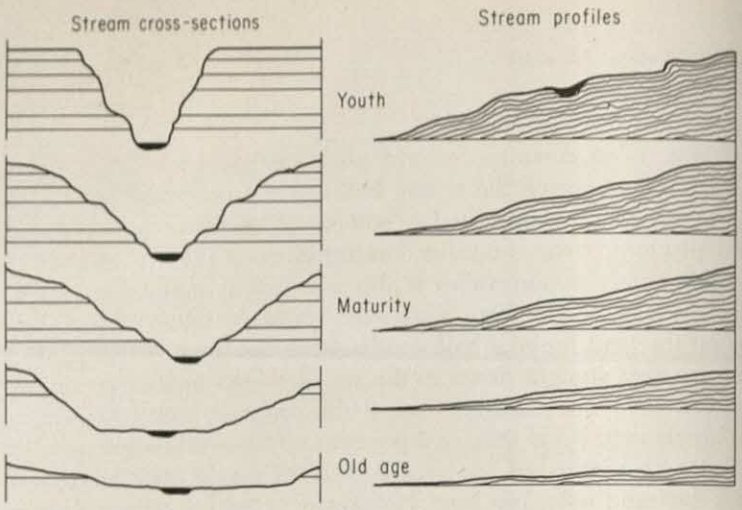


Fig. 4-2. Stages in the development of a stream: youth, maturity, and old age. (After Gilluly et al., *Principles of Geology*, San Francisco: W. H. Freeman and Co., 1951)

classic example of a stream of this type is the Colorado River as it moves through the Grand Canyon in Arizona (Fig. 4-3). Since it is in an arid climate where there is not enough water to create many tributaries, the shape of the valley is determined almost entirely by the main stream.

As a stream grows older, its steep banks are gradually rounded off and eroded into the stream; the V-shaped valley gives way to a mature valley (Figs. 4-2 and 4-3). With continued erosion, the banks become less steep and the whole valley becomes broad and shallow. The stream has then reached old age (Fig. 4-2). One of the best examples of a stream in old age is the lower Mississippi River. A map of its course below Memphis will show how it winds around in a tortuous channel through a wide, almost perfectly flat valley. In this portion of the river, it descends on its inclined plane only a few inches each mile. Although the airline distance between Cairo, Illinois, and the mouth of the river is only 650 miles, the river between these points travels over 1000 miles. Dikes, or levees, over 30 feet high have been built to keep the river from overflowing into the flat part of the valley near the stream, called the *flood plain*. Where there are no levees, the water needs to rise only a few inches to inundate large areas.

Along with the cross-sectional diagrams of a stream in its various stages, Fig. 4-2 also illustrates the changing profiles of a stream—that is, the changing elevations of the stream's bed from its source to its mouth. A stream always tends to move its source farther and farther into the highlands, thus increasing its length. As it moves downhill, it passes through rocks and soil of varying hardness and structure, and these variations produce many irregularities in its profile, such as waterfalls and lakes. The McKenzie River in Oregon, for example, starts in the upper parts of the Cascade Mountains, flows through two small lakes, drops down steep rapids, over two high falls, through more rapids, and finally emerges into a broad, flat valley. It is a stream in its youth. As the stream passes into its later stages of maturity and old age, the lakes will be drained and the falls will become eroded away until, eventually, the inclined plane over which it flows will be very smooth and gradual in its slope.

Fig. 4-3. A stream valley in youth. The Grand Canyon. (Union Pacific Railroad photo)

Fig. 4-4. A mature valley. The Green River Canyon, near the town of Green River, Wyoming. (Union Pacific Railroad photo)

Erosion is primarily the result of flowing water, and flowing water is the result of gravity. The force of gravity, therefore, is indirectly responsible for the continuous change in the face of the earth. Whether it is in the tiny rivulet that carries sand or soil a few inches in our backyard or the mighty stream that literally moves mountains of rock and soil, it is gravity that supplies the power.

4-3 Formation of Sedimentary Rocks

A flowing stream carries material in two forms—as suspended solid particles and as dissolved material. We have already seen that if a stream slows down it may unload its suspended matter. Dissolved material may also eventually be unloaded if the water evaporates to the point where the solubility (a term that will be explained in a moment) of the material is exceeded. This process is called *precipitation*.

Sedimentary rocks, as noted in Table 4-1, are rocks that have been formed from deposits of rock fragments, material precipitated from solution, or the remains of plants and animals. Streams erode away rocks and carry the fragments into lakes or the sea, where they accumulate in thick layers and are eventually transformed back into rock (Fig. 4-5). The most common types of sedimentary rock are limestone, sandstone, and shale. Salt and coal are sedimentary rocks formed from precipitation of insoluble material and organic material respectively. Let us examine the part that gravity plays in the formation of sedimentary rocks. A typical stream carries with it bits of rock, sand, and clay, and when it reaches a quiet place where the stream slows down, these

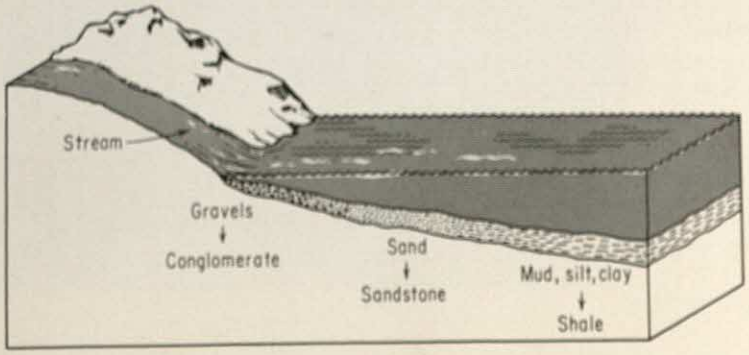


Fig. 4-5. Deposition of coarse, medium and fine particles at a shoreline.

materials settle out because the stream no longer has enough energy to suspend them. Since the density of a typical grain of sand is 2.6 g/cm³ and the density of water is only 1.0 g/cm³, the grain of sand will sink to the bottom when the speed of the stream diminishes. Eventually, most of the sand as well as the other particles reaches a lake or the sea and is deposited on the bottom, where the weight of the deposit often becomes sufficient to compact the sand particles into a hard, composite material called *sandstone*. A trace of another material, such as clay, may assist in cementing the sand grains together, but it is the downward force of gravity that is the significant factor in this process.

Most of the *limestone* in the earth's crust originates from fragments of fossils. Various kinds of hard-shelled organisms, such as clams, mollusks, and other crustacea, leave behind their shells, which are composed of calcium carbonate and have a density of approximately 2.7 g/cm³. After these organisms die, their shells may be carried by a turbulent stream into a quiet inland sea where they are dropped to the bottom. There, under increasing loads of new sediments, the calcium carbonate particles are compacted into limestone.

Shale is formed from a mixture of many kinds of suspended particles—clay, silt, and fine fragments of various kinds of rocks such as granite and limestone. All the particles that compose shale are more dense than water because they settle out when the streams that carry them flow into a large body of water. Under sufficient overload, these particles are cemented together to form shale. Almost half of the sedimentary rocks in the exposed crust of the earth are shales.

The formation of layers of mineral salt can be traced in part to the force of gravity. All streams, as they flow over rocks and through sand and gravel, collect soluble material. Even so-called freshwater streams contain traces of dissolved salts. If these streams flow into an inland sea that has no outlet, the evaporation of the water from the sea increases the concentration of the dissolved salts. In order to understand the process of salt mineral formation, we will need to examine some of the general properties of solutions. A simple solu-

TABLE 4-1 SOME OF THE PRINCIPAL SEDIMENTARY ROCKS

Name	*Composition*	*Particle Size*	*Origin*
Limestone	Calcium carbonate	...	Fossils
Shale	Silt and clay	less than .06 mm	Fragments of pre-existing rocks
Sandstone	Sand (quartz)	.06 to 2 mm	
Conglomerates	Gravels—sands and clays interspersed		
Coal	Organic matter	...	Remains of plant material
Salt	Sodium chloride, etc.	...	Precipitation from solution
Gypsum	Calcium sulfate	...	

tion is composed of two parts—a solid substance, and a liquid, in which the solid is dissolved. The ability of water to dissolve substances depends upon their nature; some substances, such as sugar, are highly soluble, and others, such as sand, are only very slightly soluble. The maximum amount of material that water can hold in solution permanently is called the *solubility* of that substance. If the solution contains less than this amount, it can pick up more material, but if it contains more than this amount as the result of the evaporation of water, it will eliminate the excess dissolved material in the process called precipitation or crystallization. If evaporation removes so much of the water in an inland lake that the remaining solution of salts exceeds their solubilities, these salts will precipitate and form crystals of solid salt. Since all salts are more dense than water, they will settle to the bottom of the sea and form a layered deposit. The least soluble salts, such as calcium sulfate, precipitate first, but eventually the most soluble salts will become crystallized as more and more water is removed by evaporation and their solubility is exceeded. In the famous Strassfurt deposits in Germany, where sedimentary salt deposits were formed in an ancient sea now well above sea level, it is possible to recover several different kinds of salts, because each one is in an almost pure layer separated from the others.

Salt layers of sodium chloride are now being formed in the Great Salt Lake, Utah, an inland sea that contains 400 million tons of salt. Since more water is being removed by evaporation than is being added by streams flowing into it, the lake every year produces 80,000 tons of salt, which crystallizes out of solution and settles under the force of gravity to the bottom. This lake is an excellent laboratory for the geologist studying the process of sedimentary rock formation, because the process is going on in it today and can be examined in detail.

In Chapter 2 we saw that geological time can be estimated by observing sedimentary rocks. The premise upon which this reckoning is based is that sedimentary rocks are laid down in horizontal layers, with the newer one being formed on top of the older ones. Gravity's role in the formation of this geological clock is obvious. It causes the streams to wash down particles of rock and soil into lakes and the oceans where they are deposited in layers that gradually get buried under such terrific weight that they are compacted into sedimentary rocks. After a particular layer of rock is formed, it may be displaced or disrupted, but we may be quite sure that the force of gravity cemented it together in a horizontal position.

4-4 Glaciers

In the snow fields of high mountains and in the polar regions, snow accumulates faster than it is removed by summer

melting. The surplus snow collects in thick beds and is compressed by its own weight into a compact mass of ice crystals known as a *glacier*. The largest glaciers in the world today cover much of Greenland and Antartica under an ice sheet that exceeds half a mile in thickness. Contrary to our usual notions about ice, a glacier is not a rigid, fixed solid. Under the force of gravity it flows much as a stream flows down an inclined plane, although at a much slower rate. Glaciers usually move only a fraction of an inch a day, but some of the larger ones on steep slopes may move as much as several feet a day. The movement of glaciers over a period of years is determined by the position of stakes driven into the top of the glacier, as well as by the movement of the rocks and pebbles that are frozen into the ice. The ice moves as though it were a stiff fluid, flowing fastest near the top and center and slowest at the edges and the bottom of the glacier. It is interesting to note that this distribution of flow is the same as that of a stream of flowing water. Geologists do not fully understand the process that makes ice flow, but they do know that the movement is powered by gravity.

There are two major types of glaciers—the valley glacier and the continental glacier. The *valley glacier* usually forms near the top of a mountain and flows downhill through a broad, sweeping channel, following either an old V-shaped valley cut by a stream of water or one that it has gouged out itself. Because of its rigidity it cannot turn sharp corners, and its path, therefore, is characterized by gentle curves. The scooping and scouring action of the glacier rounds out the bottom of the valley, making it U-shaped. Figure 4-6 shows a valley glacier in Alaska flowing slowly and picking up additional ice from several tributaries; the streaks of rock debris on its surface are telltale evidence that it is moving. In more southern latitudes, many valleys have been carved out by prehistoric glaciers that have long since melted. A fine ex-

Fig. 4-6. A long valley glacier in Alaska. Note the evidence of downslope movement shown by the debris marks where the tributaries join the main flow of the glacier. (Photo by Bradford Washburn)

ample of a glaciated U-shaped valley with gentle curves and high side walls is the Yosemite Valley in California (Fig. 4-7). The glacier removed so much rock that the tributary valleys, or *hanging valleys,* are now high above the main valley. Bridalveil Falls, in the lower right of Fig. 4-7, is at the mouth of a hanging valley.

At the head of a valley glacier, a complex action loosens and pulls away the mountainside. Snowfall is heaviest near the top of the mountain and the greatest loads of ice pile up at the head of the glacier. The downward force of gravity tends to press the glacier away from the face of the mountain, and, under the summer sun, a crevasse called a *bergschrund* often appears between them. In the winter the glacier then freezes solidly back to the cliff. This freeze-thaw cycle produces a clawing action that removes large boulders from the mountainside, and along with the scouring action of the glacier, it may hollow out the side of the mountain, creating an abrupt cliff at the head of the valley. After the glacier melts away, an amphitheater-shaped bowl, called a *cirque,* is often left behind. Sometimes the lip of the cirque is high enough to collect water and form a lake, thus producing some of our most exquisite mountain scenery (Fig. 4-8).

Since a moving glacier picks up rocks—sand, gravel, and boulders—in its downward movement, when it eventually melts and unloads this burden, it deposits the rocks in the form of *moraines.* Depending upon the size of the glacier, its load, and how it melted, a moraine may range up to miles in length and hundreds of feet in height, but usually they are much smaller. Because a glacier literally dumps its load, a moraine shows no layered structure but is a jumble of all sizes of rocks. At the lower end of a glacier, the rock and debris that have been collected by the glacier are deposited in a *terminal moraine.* A *lateral moraine* is the rock and debris that accumulate along the edges of a glacier as it moves downward. At the junction of two glaciers, two lateral moraines come together to form a *medial moraine,* which often extends down into the glacier and appears at the surface as a dirty streak, as seen in Fig. 4-6. If the climate warms up and causes the glacier to recede, *recessional moraines* are deposited in the wake of the melting glacier. Recessional and terminal moraines are quite similar and they both vary in size from a few feet to several miles in length. Sometimes they form a natural barrier to the melting water and create meltwater lakes.

The second type of glacier, the *continental glacier,* is a broad sheet of ice that extends over a very large area. In the polar regions today, continental glaciers cover large parts of Greenland, Iceland, Spitzbergen, and other Arctic islands, as well as most of Antartica.In fact, almost 10 per cent of the land area of the world is now covered with an ice sheet. During the Great Ice Ages, the last of which occurred only ten or eleven thousand years ago, an additional 20 per cent of the earth's land was covered.

Fig. 4-7. Yosemite Valley, a glaciated U-shaped valley in California.

A continental glacier moves, but not necessarily down a slope the way a valley glacier does. Many of these big broad glaciers lie on relatively flat ground, yet they move at observable rates and we naturally seek an explanation for their movement. As snow and ice pile up to great thicknesses, tremendous forces are created inside the ice sheet. At a depth of 3000 feet, for instance, there is a force of three-quarters of a ton per square inch, enough to cause the ice to be almost a fluid and to move outward in all directions from its high central point. In Greenland, the glacier moves toward the island's shore, where huge chunks break off regularly to form icebergs that float out into the North Atlantic Ocean. The Ice Age glaciers that moved out of the polar regions into North America and Europe traveled southward into temperate climates until they reached a point where they melted faster than they were replenished.

What caused these enormous glaciers to appear and disappear? Several theories have been suggested, at least five of which deserve mentioning:

1. Changes in the ratio between land and ocean areas changed the climate. Land areas change temperature more rapidly than do the oceans because of an inherent characteristic of all materials called *heat capacity*. Since water has almost five times the heat capacity of the rocks in the earth's crust, temperature variations are always more radical on land areas than on oceans, because the water greatly retards any temperature changes. If a smaller part of the earth's surface were covered with water, there would be greater temperature changes from day to day and from season to season. These temperature changes in turn would greatly affect atmospheric and oceanic currents, which are largely responsible for the character of our climate.

2. The amount of atmospheric gases and volcanic dust that shield the earth from the sun fluctuated and thus changed

Fig. 4-8. Grinnell Glacier and Lake in Glacier National Park. The glacier has retreated, leaving a magnificent cirque and lake. (National Park Service)

the earth's temperature. It is estimated that a change of only 5 to 10°C in the average temperature would be sufficient to cause continental glaciation. If a series of volcanoes spouted forth enough gas, vapor, and dust to partially block out the sun's light, the earth's temperature might drop appreciably. Man himself has hidden his cities under large quantities of smoke, vapor, dust, and fumes, and therefore it might be possible to test this theory by examining the temperature changes that are produced by man-made smog.

3. Periodic changes in the earth's orbit around the sun produced climatic variations on the earth's surface. Although it is difficult to test this theory, it is a possibility because we know that there are slight variations in the orbit of the earth even today. If the distance to the sun were at one time greater than it is now, the earth might have been cooled enough to cause glaciation.

4. Dust clouds moved in and out of the solar system and reduced the amount of the sun's heat that reached the earth. In a later chapter we will find that there are large quantities of dust and cosmic particles that occupy certain regions of the sky and that actually do obscure some of the stars at night. As they wander through space, one of these clouds might conceivably have passed near enough to act as a curtain between us and the sun.

5. The sun radiates a remarkably constant amount of heat. However, in our studies today of the energy emitted by the sun, we have found that it does give off energy at varying rates. Whether or not these variations are enough to cause great climatic changes on the earth is not known for certain, but at least this theory is one that can be partially tested by further studies of solar radiation.

Each of these theories has merit, but the best available evidence supports number two and number five. Until much more evidence is at hand, however, we cannot be sure what caused the great continental glaciers that once covered the northern part of our country.

4-5 Landslides

The principle of the inclined plane, which we have seen applied to flowing streams and moving glaciers, is similarly applicable to the downslope movement of large masses of solid materials. This movement may be slow and almost imperceptible, or it may be so rapid that the solid materials, including rock, act almost as if they were fluids. Rapid downslope movements are frequently triggered by an earthquake, although there have been major landslides that have appeared to be self-starting.

Almost every year there are serious landslides, particularly in mountainous regions, that result in heavy loss in property and lives. One of the greatest landslides in history occurred

Fig. 4-9. The Madison River slide, near Yellowstone National Park, in 1959. (U. S. Forest Service)

near the small valley town of Frank, Alberta in 1903. Ninety million tons of rock broke loose from a mountain 3100 feet above the valley floor, and this mass of rock flowed over two miles across the valley and part way up the other side. Part of the town of Frank was covered with 150 feet of rock and between sixty and one hundred people were killed. An estimated 250,000 horsepower-hours of energy, all supplied by gravity, was spent on this slide.

In more recent times, there was a similar landslide near Yellowstone Park in 1959, but this slide was started by a major earthquake. The south wall of the Madison River Canyon (Fig. 4-9) tumbled downward 1300 feet into the valley, flowed across the valley floor and dammed the river with a 300-foot barrier thus forming a new lake that submerged the forest, a highway, and a campground where several lives were lost.

Landslides ran rampant during the disastrous Chilean earthquake of 1960. This near all-time record earthquake literally crumbled dozens of mountains and dropped a 25-mile section of high country as much as a thousand feet. The coast suffered the effects of huge ocean waves set up by subterranean landslides, and these waves traveled out to sea and caused damage as far away as Tahiti and Hawaii. Property damage reached half a billion dollars and nearly two thousand perished.

In a less spectacular, but equally destructive fashion, mass movements of the soil and subsoil cause millions of dollars' worth of damage annually to roads, railroads, power and oil lines, and buildings, often in regions of only moderately rugged topography. Cities constructed on slopes are particularly susceptible to earth slides (Fig. 4-10). Rock and soil, when softened and lubricated by heavy rains, tend to move downslope, and a movement of only a few feet is enough to wreck buildings and erase property boundaries. Many cities, apparently, are planned with little or no regard for the conse-

Fig. 4-10. Downslope movement along the Pacific Coast at Santa Monica, California. (From Los Angeles *Times*)

quences of building on slopes, but in careful city planning, such factors should be considered in order to avoid injuries to property and to lives.

4-6 Variations in Gravity

In Chapter 3 we studied Newton's law of gravitation which, as summarized in Equation 3.8, says:

$$F = G\frac{m_1 m_2}{r^2}$$

This equation tells us that the force of gravity between two bodies is proportional to the masses of the two bodies and inversely proportional to the square of the distance between them. When we weigh an object, say an iron ball, we measure the force of gravity on the ball. If we moved the ball over various parts of the earth's surface, however, we would find that the weight of the ball changes slightly. Since the mass of the ball remains the same, we must assume that either the effective mass of the earth changes or the distance between the ball and the center of the earth changes. Actually, both these alternatives cause the ball's weight to change.

Let us consider first the change in distance. The greatest variation in the distance between the earth's surface and its center is produced by the bulge at the earth's equator, which results from the rotation of the earth upon its axis. Since the poles are about 13 miles nearer the center of the earth than the equator is, our iron ball weighs about 0.4 per cent more at the north pole than it does at the equator. A 200-lb man in New York City would weigh 199.5 lb at the equator, 200.8 lb at the north pole, and only 199.86 lb in Denver, Colorado, which, although it is on the same latitude as New York, is a mile above sea level. If our man were to board a space ship and fly to an altitude of 4000 miles above the earth's surface, he would be twice the earth's radius from the center of the earth and his weight would be only 50 lb.

Variations in the effective mass of the earth are not so obvious as the variations in distance discussed above. We found in Chapter 3 that the earth's mass can be measured and except for the loss of hydrogen and the gain of meteorites its total mass remains practically unchanged. The force of gravity, however, varies at points that are only short distances apart, even when we know that the points are the same distance from the center of the earth. We must, then, see if the cause of this discrepancy lies in the composition of the earth. The earth is composed of many different materials, and each one has a different density. The density of water, for instance, is 1.0 g/cm^3; the density of limestone, 2.7 g/cm^3; and of sandstone, 2.6 g/cm^3. Since an object weighs less on the ocean than it weighs on land, geologists have concluded that, since the total mass of the earth does not change, the mass of that portion of the earth near the object must affect the object's

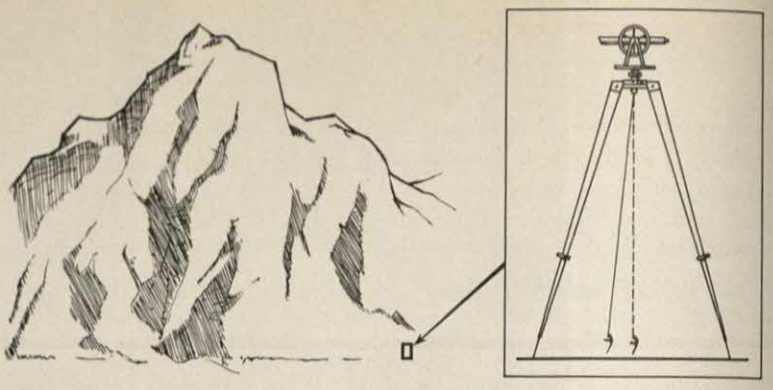


Fig. 4-11. The direction of the gravity vector displaced by a nearby mountain.

weight. The fact that the force of gravity is stronger over a dense material is exploited in the search for new deposits of ores. Fortunately, the ores of many metals are denser than most of the earth's crust. Galena, an ore of lead, has the very high density of 7.5 g/cm^3, and barite, an ore of barium, has a density of 4.6 g/cm^3. An object weighed near a large deposit of these minerals, therefore, would weigh more than it does normally. Gravity surveys, which measure the magnitude of the force of gravity, are one of the important tools of the modern geologist and prospector and have been responsible for locating several productive ore fields.

Ordinarily, the force of gravity seems to act straight downward at the surface of the earth. So dependable is this aspect of gravity that the engineer relies on his plumb line to find the vertical when he builds a skyscraper. But the assumption that the force of gravity is always toward the center of the earth has been found to be inaccurate. Not only would we expect the rotation of the earth to affect the direction of the force of gravity, but on account of some of the variations in the earth's composition cited above, the direction of a plumb line may lead to difficulties. For example, when triangulation surveys were made with a regular plum bob near the Himalayas, a serious error was made by assuming the plumb line to be exactly vertical. The presence of the high mountains nearby (Fig. 4-11) exerted enough gravitational force upon the plumb bob to pull it slightly away from the vertical, and this deviation caused surveying errors that later had to be corrected by other means, one of which involves using stars as points of reference.

These variations in the earth's gravity help us understand the forces that produce the 12-mile maximum relief—that is, the distance from the deepest part of the ocean to the top of Mt. Everest—on the earth's surface. We will now investigate the gravitational principle behind these differences in the earth's relief.

4-7 Isostasy and Archimedes' Principle

If we place two corks—one flat and thin, and the other tall and narrow—in a dish of water, we find that the tall cork rises much higher above the surface of the water than

does the thin cork (Fig. 4-12). Furthermore, if we float two corks of exactly the same size and shape, but of different grades of cork—one more dense than the other—the denser cork will sink into the water farther than the lighter one. These corks illustrate the law known as *Archimedes' principle*—named after the Greek scientist who discovered it in the third century B.C.—which says that *when a body is immersed in a liquid, it is buoyed up by a force equal to the weight of the displaced liquid.* If the body is less dense than the liquid, it floats, as do the two corks in Fig. 4-12.

If we cut off part of the tall cork in Fig. 4-12 and place it on the thin cork, the tall cork will rise slightly and the thin one will sink accordingly. Or, with our two corks of different density, if we cut a slice off the top of the lighter cork and place it on top of the heavier cork, the lighter cork will rise and the heavier cork will sink. These two pairs of corks illustrate the concept of isostasy, one of the fundamental principles of geology. *Isostasy* is a state of balance between the various components of the earth's crust. If this balance is disturbed by erosion, glaciation, landslides, or sedimentation, the parts of the earth's crust that are affected rise or sink in an attempt to re-establish a new state of equilibrium.

The tall cork in Fig. 4-12 is analogous to high mountains and the flat cork to lowlands. As streams remove material from the mountains and deposit it on the lowlands, the weight of the mountains is reduced, and they rise slightly. Consequently, the lowlands, as more and more sediments pour into them, are forced downward under their increasing burden. Some geologists, however, think isostasy is more like the two corks of different density; they hold that the mountains are composed of less dense material than the lowlands and that the rock that is removed from the mountains by erosion is converted into a more dense rock in the valley. According to this latter concept, the bottoms of the mountains are nearly the same level as the bottoms of the lowlands, whereas in the former theory the mountains' roots extend lower than the bottoms of the lowlands.

Actual studies of the earth's surface have revealed that these theories are plausible. The Central Valley of California, for instance, seems to have always been at approximately the same elevation, yet new sediments have been constantly added to it during the millions of years of geological history. Each new layer of sediment has increased the load on the earth's crust and has depressed the many horizontal layers of sedimentary rock that now underlie the valley. Figure 4-5 shows how layers of sediment might accumulate in the shallow part of the ocean. The constant depth of the water indicates that the new loads of sediment force the lower layers still lower. At the same time, the land areas around this shallow part of the sea rise as they lose weight through erosion. In the same way, the mountains on either side of the Central Valley of California are also rising.

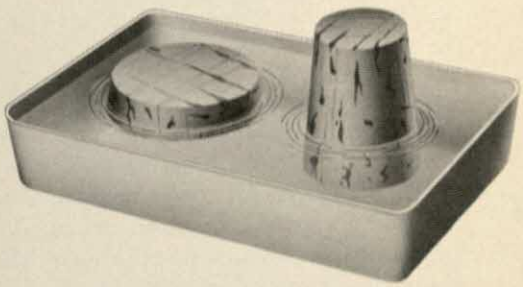


Fig. 4-12. Two corks of different shape floating on water.

The principle of isostasy is based upon the existence of a subterranean fluid under the continents that has a greater density than the continental mass and that yields to the weight of its overlying burden. This fluid is not a liquid in the ordinary sense; it is a very dense substance that becomes viscous only under the tremendous pressures of the rock layers above. It is probably ordinary rock, which in large masses, as we saw in the Frank, Alberta slide, is capable of acting like a fluid. Although the principle of isostasy is not fully substantiated, it does explain why there can be mountains even though the process of erosion would appear to wear them down. The altitude of some of our present peaks, such as those in the Himalayas, is even becoming higher, indicating that during recent geological periods the principle of isostasy has been producing higher mountains. The earth's crust has probably been rough during much of the past, and there is probably more relief now than ever.

4-8 The Tides

The ebb and flow of the tides have been observed since ancient times. Twice a day the ocean level rises and falls, and, ever since the discovery of the universal law of gravity, this pulsating motion has been attributed to the gravitational forces between the earth and the moon. The method by which this force of gravity causes the tides is illustrated in Fig. 4-13.

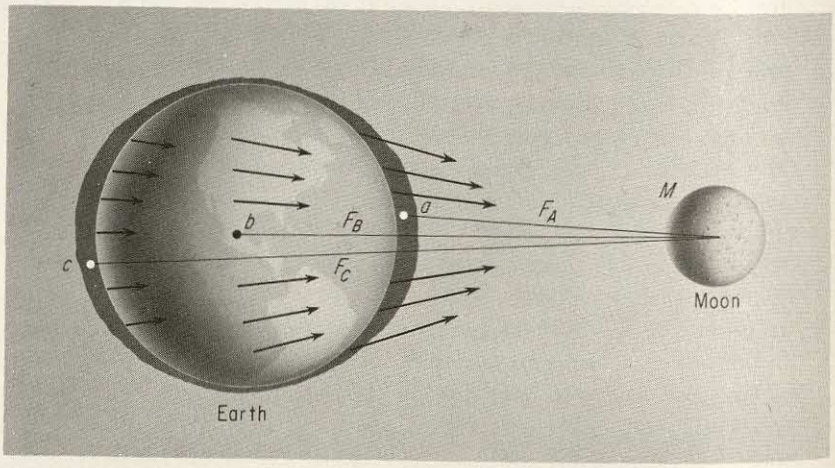


Fig. 4-13. Gravitational force of the moon pulling on the oceans. (From Sverdrup, Fleming, and Johnson, *The Oceans* (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1942, and others)

As we know, the force of gravity between two bodies can be calculated from the equation:

$$F = G\frac{Mm}{r^2}$$

If we let M be the mass of the moon (7.3×10^{21} kg), m a metric ton of mass (10^3 kg)—either of water in the ocean or material at the center of the earth—and r the distance from the center of the moon to the center of the earth (3.91×10^9 m), we can calculate the three different forces of gravity, F_A,

F_B, and F_C, in Fig. 4-13. The forces on a metric ton of mass at *a*, *b*, and *c* respectively are:

$$F_A = 6.67 \times 10^{-11} \frac{(7.3 \times 10^{21}\ \text{kg})(10^3\ \text{kg})}{(3.85 \times 10^9\ \text{m})^2}$$

$$= 3.3 \times 10^{-5}\ \text{newtons}$$

$$F_B = 6.67 \times 10^{-11} \frac{(7.3 \times 10^{21}\ \text{kg})(10^3\ \text{kg})}{(3.91 \times 10^9\ \text{m})^2}$$

$$= 3.2 \times 10^{-5}\ \text{newtons}$$

$$F_C = 6.67 \times 10^{-11} \frac{(7.3 \times 10^{21}\ \text{kg})(10^3\ \text{kg})}{(3.97 \times 10^9\ \text{m})^2}$$

$$= 3.1 \times 10^{-5}\ \text{newtons}$$

From these calculations, we can readily see that the moon exerts noticeably more gravitational pull on a body on the side of the earth facing it than it does on the side away from it.

If the mass at *a* is water, it will tend to bulge outward toward the moon. The rock crust of the earth may also bulge outward, but it will not bulge very far because it is much more rigid. At *c*, the mass of water will be attracted less than the same mass at *b*, so the earth will be pulled away from the water at *c*, thus producing a bulge on the opposite side of the earth from *a*. The water in the oceans is held to the earth by the earth's gravity, and at the same time it is drawn away from the earth by the moon's gravity on the other side. The resultant force of these gravitational forces acting on *a* and *c* are shown by the vectors in Fig. 4-13.

In addition to the moon's gravitational force, the sun's gravity also attracts our ocean waters. Although the sun is much farther away from us than the moon, its great mass produces a gravitational pull on the tides that is 46 per cent as powerful as that of the moon. When the moon and the sun are on a line, as in positions *A*, *B*, and *E* in Fig. 4-14, the grav-

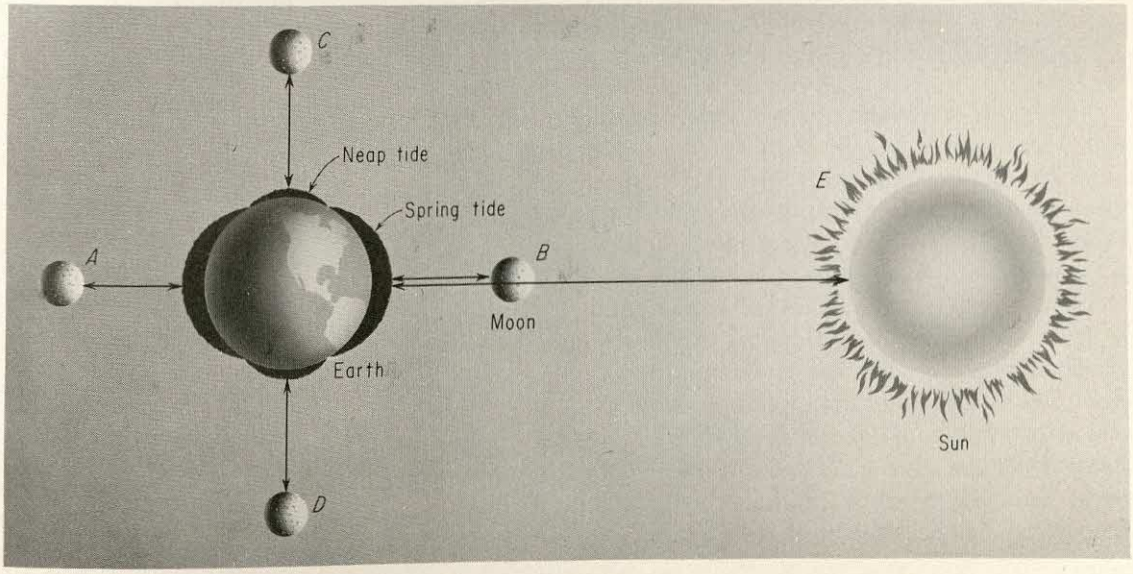


Fig. 4-14. The combined effect of the sun and the moon on the earth's tides. (From Sverdrup, Fleming, and Johnson, *The Oceans*, Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1942, and others)

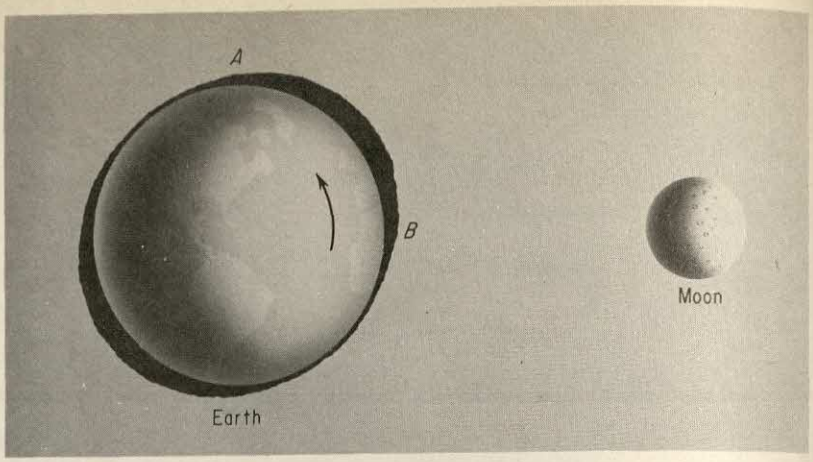


Fig. 4-15. The rotation of the earth into the tidal bulges makes the tides. (From Sverdrup, Fleming, and Johnson, *The Oceans*, Englewood Cliffs, N. J.: Prentice-Hall Inc., 1942, and others)

itational forces of the moon and the sun reinforce each other and produce the high tides called *spring tides*. When the moon and the sun are at right angles to each other as in positions *C*, *D*, and *E*, they partially neutralize each other's effect upon the tides; the resulting small tides are called *neap tides*.

As the earth rotates on its axis, the bulge in the ocean waters tends to remain fixed with respect to the moon. This means that a person standing on an ocean shore line at *B* in Fig. 4-15 would see the tide rise, because that part of the earth is moving into the bulge. A person at *A* will experience a falling tide because the earth there is turning away from the bulge. The same process occurs on the back side of the earth; thus there are always two regions of high tides and two of low tides on the earth at any one time.

In Chapter 2 we found that the earth's rate of rotation is gradually slowing down because the tides act as a brake on the earth's spinning. Most of the braking friction is between the oceans and their shallow beds near the shore lines. The tidal bulges, in short, act as two huge brake-bands attached to the opposite sides of the earth. This tidal friction amounts to 1.3×10^{15} pounds of force, enough to slow down the earth's rate of rotation one second in 100,000 years.

4-9 Gravity and the Atmosphere

The earth is surrounded by a gaseous envelope, called the *atmosphere*, which is kept from moving out into space by the earth's gravity. Since the gases that compose the atmosphere are compressible, the density of the atmosphere decreases rapidly as the distance from the earth increases. In fact, one-half of the atmosphere lies within about 3.5 miles of the surface, and 99 per cent lies below an altitude of 20 miles. In their observations of meteorites and the aurora borealis (Northern Lights), scientists have found traces of the atmosphere 500 to 750 miles above the earth. If the atmosphere were not compressible and if all of it had the density it has at sea level, the 11.4×10^{18} lb weight of the atmosphere would form a layer only 4.9 miles high. We will leave the detailed study of the structure of the atmosphere for a later chapter.

Our smaller neighbor, the moon, has no atmosphere surrounding it. The reason for this lack can be found in the difference between the gravitational pull on the surfaces of the moon and the earth. We can calculate this difference by applying the gravitational law in which M_e is the mass of the earth, M_m the mass of the moon, m a unit mass on the surface of the earth and on the surface of the moon, and r_e and r_m the radii of the earth and the moon respectively. Thus, the force of gravity acting on the mass m at the surface of the earth may be expressed:

$$F_e = G\frac{M_e m}{r_e^2}$$

The same mass m is attracted to the surface of the moon with a force:

$$F_m = G\frac{M_m m}{r_m^2}$$

Dividing the first expression by the second, we get:

$$\frac{F_e}{F_m} = \frac{G M_e m}{r_e^2} \times \frac{r_m^2}{G M_m m} = \frac{M_e r_m^2}{r_e^2 M_m}$$

The mass of the earth is 6×10^{24} kg, and its radius is 6.4×10^6 m. The mass of the moon is 7.3×10^{21} kg, and its radius is 1.8×10^6 m. Thus, the relationship between the force of gravity on the earth and that on the moon can be calculated by substituting these quantities in the equation above:

$$\frac{F_e}{F_m} = \frac{6 \times 10^{24}\ \text{kg} \times (1.8 \times 10^6\ m)^2}{(6.4 \times 10^6\ m)^2 \times 7.3 \times 10^{21}\ \text{kg}} = 6.5$$

The force of gravity on the earth, therefore, is 6.5 times as great as that on the moon. A liter of air on the earth weighs 1.3 g/l, but on the moon it would weigh about one-sixth that much. The moon's gravity, however, is not great enough to hold any atmosphere, so any gas that would happen to collect on it would be diffused into space.

Many of the other planets in the solar system have atmospheres about them. The masses of Jupiter, Venus, and Mars are large enough to create a force of gravity capable of holding the gases that comprise their atmospheres. The atmospheres of these planets, however, are unlike the earth's in that they are composed of gases other than nitrogen and oxygen: The sun's atmosphere contains large amounts of hydrogen and helium; the atmosphere of Jupiter contains methane, ammonia, hydrogen, and helium; the atmospheres around Venus and Mars contain a considerable quantity of carbon dioxide.

The earth's gravity is not only sufficient to hold an atmosphere but is also powerful enough to give the atmosphere a downward force of 14.7 pounds per square inch at sea level. In other words, above each square inch of surface at sea level there is 14.7 pounds of air. Since this weight is a force per unit area, it is called *pressure*. Some gases are denser than air and therefore exert more pressure than this. For instance,

a dense gas, such as carbon dioxide—whose density is 1.9 g/l compared to the 1.3 g/l density of air—tends to lie near the ground and to flow like water into low-lying areas. Even though gases ordinarily mix with each other, if a large quantity of gas is released into the air, a period of time elapses before it becomes dispersed. If the released gas is more dense than air, then gravity will hold it near the ground where it displaces the air. For example, in the Valley of Death on the island of Java, enough carbon dioxide has accumulated from gas wells in the vicinity to make the valley a death trap for both animals and human beings. Occasionally, tunnels, mines, and even buildings may accumulate enough carbon dioxide to make them unsafe for workers.

In our ordinary experience, we know that warm air rises and that cool air descends. The reason for this is that cool air is more dense than warm air, a fact that will become clearer when we examine the structure of gases in a later chapter, where we will find that as a gas warms it expands and decreases its weight per unit volume. For instance, air at 32°F. is 1.14 times as dense as air at 100°F. When cool air accumulates in upper mountainous regions, since it is heavier than warm air, it flows downhill, displacing the warmer, lighter air. If fairly steady and powerful, these movements of cool air are called *katabatic winds* and may reach velocities of 100 to 200 miles per hour. They occur in the Santa Ana Canyon of southern California and are quite common in the snow-covered mountains of Greenland. In the Antartic, a mysterious phenomenon that puzzled scientists from the time of the first explorations has been found to be a katabatic wind. The weather is usually calm on the Ross Sea, but a short distance away, off the Adélie coast, winds up to 200 miles per hour blow all winter long. Recent surveys have revealed a wide, sloping trough running from the South Pole to the Adélie coast. During the winter, the bitterly cold air at the pole spills down this trough, accelerates through the 1500 miles of the natural inclined plane, and lashes out over the sea in a howling gale.

4-10 Summary

Gravity is the driving force behind many of nature's processes. Although raindrops descend in a free fall and cause some displacement of the soil, water does much more damage when flowing in streams down an inclined plane. Streams wash away rock and soil in the process called erosion and then deposit them in a delta or farther out into a lake or the sea. As these deposits accumulate on the bottom of a lake or the ocean, their tremendous weight often compresses them into sedimentary rocks.

Both kinds of glaciers—valley and continental—move under the force of gravity and are powerful agents of erosion. Although several theories have been advanced to explain the

great continental glaciers of the four Ice Ages, none of them is accepted without reservations because our information is inadequate and because they cannot be tested experimentally. Gravity also forces earth downhill, sometimes in swift cascading landslides and sometimes in slow movements of the soil.

The force of gravity varies over the earth's surface; it is greatest at low elevations, at the earth's poles, and in regions where the earth's crust is very dense. The direction of the gravity vector may not always point straight downward, because it is influenced by the mass of nearby mountains.

The continents "float" on a very dense "fluid," and their mountains and lowlands tend to maintain an equilibrium called isostasy.

The water in the oceans is pulled outward by gravitational force toward the moon, and to some extent toward the sun, and at the point where the earth moves into one of these bulges, there is a high tide. The friction of the tides with the earth's surface is slowing down the earth's rate of rotation one second in 100,000 years.

The earth's atmospere is kept on the planet by the force of gravity. The moon's gravity, however, is not strong enough to hold an atmosphere. Dense gases and cool air sink to the ground, sometimes causing gaseous death traps and fearsome katabatic winds.

EXERCISES

A. KEY TERMS TO REMEMBER

alluvial fan
Archimedes' principle
atmosphere
bergschrund
cirque
continental glacier
delta
erosion
flood plain
hanging valley
isostasy
katabatic wind
lateral moraine
limestone
mature valley
moraine
neap tide
old-age valley
sandstone
sedimentary rocks
shale
spring tide
terminal moraine
U-shaped valley
V-shaped valley
valley glacier
youthful valley

B. QUESTIONS ABOUT GRAVITY IN NATURAL PROCESSES

1. Since stream erosion is caused by water flowing down an inclined plane, suggest several ways by which erosion of the soil might be controlled or reduced.

2. Suggest several problems that should be considerd in building a city, a highway, or a railroad in a region of potential landslides.

3. Explain why sedimentary rocks, when first formed, are always in horizontal layers.

4. How can you tell from an airplane whether or not a continental glacier has covered a particular section of land?

5. What information would you need in order to prove or disprove the validity of each of the five explanations that have been offered for the appearance and disappearance of the continental glaciers?

6. If you were buying gold on top of a mountain and selling it at sea level, would it make any difference whether you used a spring scale or a balance to measure the metal?

7. Does the isostasy principle affect the estimates of how much dirt and rock should be excavated to accommodate the foundation of a tall skyscraper?

8. Suggest a method of harnessing the tides to do useful work, keeping in mind that the force of gravity is the primary source of this energy.

9. In a region of low hills, where would you expect to find the warmest air at night, on the hilltops or in the valleys?

10. The density of a war gas, phosgene, is almost 3.5 times as dense as air. If you didn't have a gas mask, how else might you protect yourself from this gas?

11. Which of the following are the effects of erosion by a valley glacier: (a) U-shaped valleys, (b) cirques, (c) hanging tributaries, (d) moraines.

12. In what other ways would the energy of a flowing stream be dissipated besides friction with its bed and banks?

13. Which of the following features in sedimentary rocks are the result of sedimentation: (a) stratification, (b) sorting, (c) crossbedding, (d) caves.

14. Which of the following may consolidate sediments into sedimentary rocks: (a) compaction, (b) cementation, (c) solution, (d) recrystallization.

15. Which of the following properties of carbon dioxide contributed to the deadly nature of the "Valley of Death"? (a) does not support combustion, (b) colorless, (c) odorless, (d) more dense than air.

16. Explain why a liter of gas released on the moon may eventually be lost to space.

C. PROBLEMS ABOUT GRAVITY IN NATURAL PROCESSES

1. If an earthmoving machine is rated at 200 horsepower, how many of them, working day and night, would it take to produce the power generated by all the streams and rivers in the world? Ans. 6,200,000.

2. Select a valley with which you are acquainted, or examine one near your home, and draw a rough cross section and profile of it. In what stage of development is it?

3. By means of two or three diagrams, show how the head of a valley glacier can transform a smooth face of a mountainside into a cirque.

4. In your community, find several examples where erosion has been chiefly responsible for the appearance of the landscape.

5. By diagram, show what the cross section of a stream valley would be if it cut its way through to bedrock in a flat region composed of the sedimentary rocks shown in Fig. 4-5. Assume that shale is relatively soft, sandstone intermediate, and limestone hardest and most resistant to erosion.

6. At what points on the earth's surface, and under what conditions would a plumb line point exactly toward the center of the earth?

Five

Force and Motion

In conversation we commonly use the word "force" when speaking of the force of the wind, the force of ocean waves, the force of gravity, and the force of a bat as it strikes a ball. All these things do, indeed, exert a force, but when we try to formulate a scientific definition of force, we encounter a difficulty. Ordinarily, we do not see a force itself; we see the wind turn a windmill, the ocean waves cut away a shore line, and a well-swung bat send a ball over the outfield fence. *Force*, then, is defined as any influence that causes, or tends to cause a change in the motion of an object. In other words, when a force is applied to a body, it accelerates that body by changing, through a push or a pull, either its direction or its rate of movement.

Force is related to motion because force is necessary to produce motion. In spite of a magician's tricks, a stationary object never moves unless some kind of force is applied to it. It is the movement of the wind, the movement of the ocean waves, and the movement of the bat that provides the force in the above examples. We could not analyze the force of gravity if we were prohibited from studying the motion it produces in, for example, a falling apple or a ball rolling downhill. In addition to motion, force is also related to mass. For instance, if a gnat strikes our face at a velocity of 30 mi/hr we barely feel the impact. But if we meet a brickbat face-on at 30 mi/hr it is quite another matter. The brickbat obviously exerts more force on us because it has a much greater mass than the gnat. These two factors—motion and mass—are of primary concern in understanding the mechanical nature of force. The basic relationships between force, motion, and mass were concisely explained by Newton in his *Princi-*

pia. After a series of brilliant experiments and a keen analysis of the results of these experiments, as well as a study of the motion of the moon and the planets, Newton proposed a set of laws or principles that govern forces and motion. Although much of his thinking was done with the older forms of mathematics, he invented a new kind—the calculus—in order to fully describe these laws.

In this chapter we are going to examine Newton's three laws of motion and some of the applications of these laws to common, everyday situations. Galileo came very close to formulating one of these laws, but it was not until half a century after his death that Newton expressed these concepts in terms that could be applied to many practical situations. So basic are these three laws that if Newton had contributed nothing else to science, he would still be honored as a giant among the scientists of all time.

5-1 Newton's First Law

Prior to Newton's time, the motion of an object was believed to be the result of a mysterious force that continually pushed the object. The moon, for example, was supposed to move in its orbit around the earth because something was always pushing it. If this strange force were to cease, the moon would stop, in much the same way as a ball rolling on a smooth surface eventually stops unless we give it a push occasionally. But Newton, after much study, deep thinking, and experimenting, realized that the ball slows down because of friction, and that if the moon experiences no friction it will continue moving without being pushed. This conclusion led Newton to the formulation of his first law of motion: *A body at rest or in motion will remain at rest or in motion in a straight line unless acted upon by some external force*. In order to fully understand the meaning of this law we must first examine a property inherent in all matter—inertia.

Suppose that you are a passenger on a bus who unfortunately must stand and that the bus is traveling along a highway at 50 mi/hr. If the bus suddenly slows down, you, along with all the other standing patrons, experience a sudden forward thrust—it feels as though you might all be thrown into the front end of the bus. What is the cause of this forward thrust, or is it a forward thrust? Actually, when the bus slows down you and all the passengers continue to move forward at 50 mi/hr, unless you grab a strap or the hand rail. You are experiencing that quality of all matter called *inertia*, which is an object's resistance to being set in motion or to a change in its motion. The word "inertia" is derived from the Latin word meaning "laziness," and it aptly describes the inherent tendency of an object to resist any change of motion. The amount of inertia that an object possesses is proportional to the amount of force required to slow it down, speed it up, or change its direction.

It is quite difficult to make precise experiments that demonstrate the first law because the presence of friction and of the force of gravity prevent an object from traveling at a constant speed in an exactly straight line. One of the best proofs of the law is found in the movement of the celestial bodies, which encounter practically no friction in their travels. The only perceptible forces acting on the planets are the forces of gravity that hold them in their orbits around the sun, and, as far as we know, the planets have been moving at a precisely constant rate ever since they have been observed by man. The behavior of these bodies gives ample proof that an object would move in a straight line if it were not acted upon by external forces, thus confirming the first law, which eliminates the necessity of a mysterious force to preserve the movement of a moving object.

5.5 Newton's Second Law

What happens to the motion of a body when it is acted upon by a force? We know that a force can change the motion of an object, but Newton was interested in finding a precise relationship between force, motion, and mass, and he answered this question in the statement of the second law, which says: *The acceleration (or change in motion) which a force imparts to a given mass is directly proportional to the force and inversely proportional to the mass.* In mathematical terms the second law has a double relationship—a direct and an inverse proportion. A simple equation that expresses this relationship is:

$$a = \frac{F}{m}$$

or, by rearranging these three terms, we can say:

$$F = ma$$

If we choose suitable units for the quantities in the three terms in this relationship, we can write:

$$F = ma \qquad (5.1)$$

For instance, if we express the mass in kilograms and the acceleration in meters per second per second, then the force is in newtons, a basic unit of force in the metric system. A newton then is the force required to accelerate a kilogram mass at 1 m/sec². We can use this as a means of defining a kilogram: *the mass that is accelerated at 1 m/sec² by a force of 1 newton.* The most common example of the second law is the force of gravity acting upon a mass, a phenomenon we have already examined in Chapter 1. The force of gravity upon an object causes it to be pulled toward the earth's center with a force we call its weight (Fig. 5-1). For this special case we can rewrite Equation 5-1 by substituting weight for force and the acceleration due to gravity, g for a

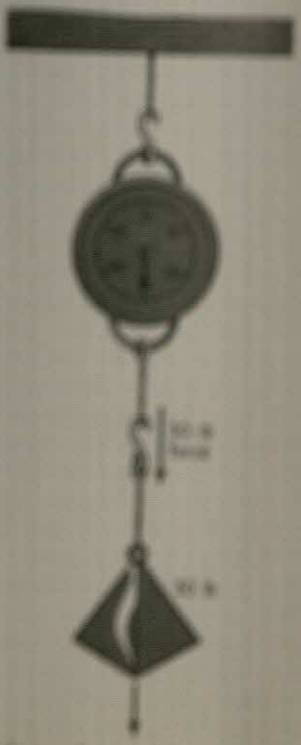


Fig. 5-1 Weight is a downward force.

Then

$$W = m \times g$$

The direction of this force, W, can be found if we recall that gravitational acceleration is a vector pointing toward the earth's center and that mass is only a scalar. As you know, a vector has magnitude and direction, while a scalar has only magnitude, and thus when we multiply a vector by a scalar we merely change the magnitude of the vector but not its direction. The direction of the force (weight) of an object at the earth's surface, therefore, is directly downward. On a larger scale, we can find that the direction of the vector representing the force of gravitation between the earth and the moon lies exactly on the line that passes from the center of the earth to the center of the moon. We shall examine in more detail these forces in the solar system in Chapter 7.

An experiment that illustrates Newton's second law is shown in Fig. 3-5, where a car is being accelerated by a constant force F supplied by a cord attached to a spring scale which can move so that the same force is always exerted on the car. If the car is originally at rest, the force will start it rolling and accelerate it along the board. A larger force would accelerate the car more rapidly, but if the car were much heavier, it would be accelerated more slowly. If we inspect Equation 3-1 we will find all these relationships. The applied force is directly proportional to the mass of the car, the force is directly proportional to the acceleration of the car, and the mass of the car is inversely proportional to the acceleration of the car.

In order to illustrate the second law, let us work an example problem. If we assume there is no friction, how many newtons of force are required to give a 1000-kg car an acceleration of 2 m/sec^2? (This car would weigh about 2200 lb and would be accelerated to 80 mi/hr in about 11 seconds.) Solution: by substituting these quantities into Equation 3-1, we get

$$F = ma = 1000 \text{ kg} \times 2 \text{ m/sec}^2$$
$$F = 2000 \text{ kg m/sec}^2 = 2000 \text{ newtons}$$

In the solution of this problem, note that when the mass in kilograms is multiplied by acceleration in m/sec^2, the units of the product are kg m/sec^2—precisely the units of a newton of force. Before considering Newton's third law, let us examine two implications of the second law—momentum and work—that involve common and fundamental characteristics of force and motion.

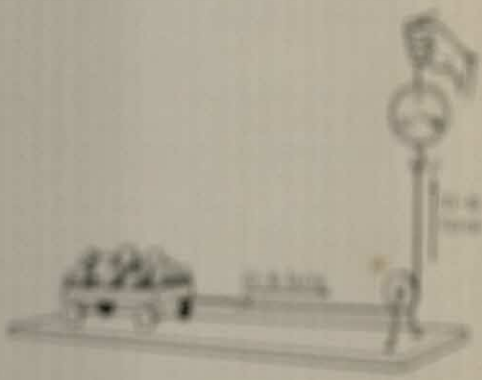


Fig. 3-5 The acceleration of a car when acted upon by a force.

3-5 Momentum

In our common experience we are aware of the fact that there is a basic relationship between mass and velocity. For instance, we know that both a bullet fired from a gun and a

glacier sliding slowly down a mountainside are destructive —the former being a small mass with a large velocity and the latter a large mass with a very small velocity. But what if the bullet traveled slowly and the glacier rapidly? This would obviously make a vast difference in their effects. The combination of the mass and the velocity of an object that concerns us here is the property of matter known as momentum.

Momentum (its symbol is p) is the product of the mass of an object and its velocity, or:

$$p = m \times v \tag{5.2}$$

All moving objects have momentum. If the mass of a moving object is doubled and if the object maintains the same velocity, its momentum is also doubled. Thus, a 200-ton locomotive moving at 60 mi/hr has twice the momentum of a 100-ton locomotive moving at the same velocity. The little car in Fig. 5-2 has an initial momentum of mv_0; after the force has acted upon it for a time its momentum is mv. The change in momentum, $mv - mv_0$, is proportional to the force acting on the car and to the time it acted.

As a very simple illustration of the way some scientists might examine momentum, let us look at the units involved.

$$F = ma \quad \text{or newtons} = \text{kg}\frac{m}{\text{sec}^2}$$

$$p = mv \quad \text{or kg}\frac{m}{\text{sec}}$$

Now comparing these two, the only difference seems to be in the denominator, with newtons having sec^2 while momentum has only sec. If we multiply newtons by time, we get the same units as momentum and in a very simple way we will have derived, theoretically, a new formula: $F = ma$; multiplying by t, $Ft = mat$, but from Equation 3.3, $v = at$; therefore $Ft = mv$ or newton sec $= \text{kg}\frac{m}{\text{sec}}$. This involves the application of a force to an object during a time interval and is called *impulse*. Unless the object is at either an initial or final velocity of zero, the more useful form of the impulse equation is

$$F(t_2 - t_1) = mv_2 - mv_1 \tag{5.3}$$

Thus, impulse is the change in momentum.

Impulse and momentum are very important quantities. A car traveling down the road and the people in the car each have a certain momentum. If this car and the people in it are to be slowed or stopped, an impulse has to be provided just equal to the change in momentum. We can have a large force for a very short time or a smaller force for a much longer time. Since large forces can be very destructive, crumple fenders, and fracture skulls, we had best lengthen the stopping time and thereby decrease the force needed to make the change of momentum. In case of an accident, seat

belts tend to increase the stopping time for the passengers and consequently, less potentially destructive forces are applied to their bodies.

A hammer hitting a nail, a bat hitting a ball, a club striking a golf ball, a car hitting a power pole, or a person's head hitting the windshield—all these contacts are examples of impulse and change in momentum. The longer the time this change in momentum takes place, the smaller is the force involved. From Equation 5.2 we see that the momentum of an object is proportional to both the mass and the velocity of an object. For this reason it would be one thing to swing a bat at a baseball traveling 60 mi/hr and quite another thing to swing a bat at a locomotive traveling 60 mi/hr. The impulse transferred to the locomotive from the bat might be just the same as that transferred to the ball but because of the great mass of the locomotive, it would change the velocity of the locomotive an insignificant amount, while it might be enough to drive the ball over the fence.

5-4 Work

Another aspect of the second law is the physical concept of work. In spite of the usual meaning of work—bookwork, housework, and so on—work, as briefly mentioned in Chapter 2, is the measure of a force having moved through some distance. *Work* is most simply defined as force times distance; it is the result of a force acting through a certain distance. Let us consider an example so that we are sure we understand this concept.

If a book with a mass of 1 kg is lifted 1 m, what work is performed? According to Newton's second law, the force required to lift an object (its weight) is its mass times the acceleration of gravity (Equation 3.10), or:

$$\text{weight} = \text{mass} \times \text{acceleration} = mg = 1\ \text{kg} \times 9.8\ \text{m/sec}^2 = 9.8\ \text{newtons}$$

If we lift the book 1 m, then the work is found from the relationship:

$$\text{work} = \text{force} \times \text{distance} = 9.8\ \text{newtons} \times 1\ \text{meter} = 9.8\ \text{newton meters} \qquad (5.4)$$

The unit newton meter is used frequently enough to have its own name—a joule. A *joule* is a unit of work; its units are newton meters. In the English system of units, work would be the product of force in pounds and distance in feet or foot-pounds. If we had to exert 50 lb of force to move an object 20 feet, we would do 1000 ft-lb of work. While we seldom change from one system of units to another, you might find it interesting to show that 1 ft-lb of work equals 1.356 joules. In later chapters we will see that energy may have many forms; in electrical units, 1 joule = 1 watt second; in heat units, 1 joule = 0.23889 calories. Scientists generally do not memorize conversion units such as those just given, but they

are aware that work may have several forms and corresponding units. For the exact numerical equivalents between these forms they will most likely use a table of conversion factors.

Sometimes it is not enough just to do a certain amount of work, but the work must be done at a certain rate. To describe how fast work is done, we use the term *power*. One joule of work per second is called a watt, while 550 ft-lb per second is a horsepower. We can write this in the form of a formula:

$$P = \frac{\text{work}}{\text{time}} (\text{joules/sec or ft-lb/sec or watts or horsepower}) \quad (5.5)$$

Perhaps an example will help.

Suppose a jet airplane weighing 150 tons (300,000 lb) can climb 1000 ft/minute. What minimum horsepower would this represent? Work done in one minute of time:

$$\text{force} \times \text{distance} = 300{,}000 \text{ lb} \times 1000 \text{ ft} = 300{,}000{,}000 \text{ ft-lb}$$

Power:
$$\frac{\text{work}}{\text{time}} = \frac{300{,}000{,}000 \dfrac{\text{ft-lb}}{\text{min}}}{60 \dfrac{\text{sec}}{\text{min}}} = 5{,}000{,}000 \text{ ft-lb/sec}$$

Horsepower:

$$\frac{\text{ft-lb/sec}}{550 \text{ ft-lb/sec/hp}} = \frac{5{,}000{,}000}{550} = 9{,}090 \text{ horsepower}$$

Thus, we see that the common term *horsepower* is a unit of work performed at a certain rate, and that in designing an engine of a given horsepower we are interested in a device that can do work for us at a given rate.

5-5 Newton's Third Law

Of Newton's three laws of motion, the third one is perhaps the least understood. This third law says: *For every force there is an equal and opposite force.* There is no such thing as a single force existing all by itself. Newton illustrated this concept by pointing out that when you press a stone with your finger, your finger is also pressed by the stone. Similarly, when you stand on the floor, you exert a downward force on the floor and at the same time it pushes upward on your feet with exactly the same force. The third law tells us that a force always involves two bodies acting upon each other. The wind exerts a force upon us as we walk along a windy street, but we also exert a force upon the wind. A bat exerts a force as it strikes a ball, but the ball in turn exerts a force upon the bat. When we pull a block of wood along a smooth surface with a string, the block of wood pulls on the string in the opposite direction. If the block of wood exerted no force in the opposite direction, it would take no force to pull the block along in the first place.

The crucial point to remember here is that no force can operate independently of another. The earth's gravity attracts

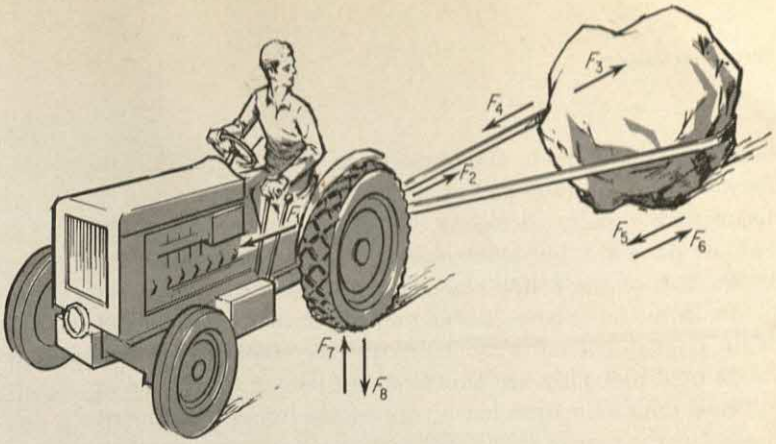


Fig. 5-3. The forces of action and reaction—an illustration of Newton's third law.

a ball to its surface, but at the same time, according to the law, the ball pulls the earth upward. The upward movement of the earth is immeasurably small because the earth's mass is so very much greater than the ball's mass. A baseball struck by a swinging bat exerts an equal force on the bat, but in the opposite direction. The force of the bat accelerates the ball in one direction, while at the same time the ball accelerates the bat in the opposite direction. In effect, this interaction slows down the velocity of the bat and changes the velocity of the ball, even though the ball's speed might remain unchanged.

To use another common example, suppose a tractor is pulling on a rock (Fig. 5-3). In this illustration, there are several forces of action and reaction. The tractor pulls the rope forward with a force F_1, while the rope pulls the tractor with an equal force F_2 in the opposite direction. The rock exerts a force F_3 on the rope, and the rope exerts an equal and opposite force F_4 on the rock. The rock drags along the ground with a friction force F_5, and the earth resists the rock with a force F_6. Force F_7 is that of the earth pushing up on the tractor, and F_8 is the weight of the tractor pushing down on the earth. If all the forces are equal, the system is in equilibrium. But if F_1 (the force of the tractor pulling forward) exceeds F_5 (the force of friction), the whole system will move forward with an acceleration that can be derived from the equation, $F_1 - F_5 = ma$. Now you can see that the forces F_7 and F_8 should be inclined to the vertical, for the tractor must push backward on the earth by an amount equal to the force with which it is pulling ahead on the rope. In the same way, the earth must push ahead on the tractor or it will slip and be unable to exert force F_1 on the rope. As the whole system moves forward, the earth itself is moved backward (though of course only slightly) because there is always a reaction to every action.

If an observer out in space were to watch a railroad train that stretched a third of the way around the globe start from rest, he would see another instance of the third law in operation (Fig. 5-4). He would see the train move forward with a certain acceleration, and at the same time he would see the earth accelerate in the opposite direction. Since the mass of the earth is so much larger than that of an ordinary train, the actual acceleration of the earth when a train starts up is

Fig. 5-4. As a train accelerates forward the earth is accelerated in the opposite direction according to Newton's third law.

extremely small. The inertia of the earth is vastly greater than the inertia of the train, and because of the great difference in masses, we usually disregard the effect of an accelerating train on the earth, but nevertheless it is there. A very modern application of the action-reaction principle is in the jet airplane. In the jet engine, fuel and air are drawn into the front of the engine where the fuel is burned very rapidly and then the burned fuel gases are blown out the rear of the engine at a tremendous velocity. That is, the engine forces the exhaust gases out the rear of the engine at a much greater rate than the air is taken in at the front. The reaction to this force is a forward force that actually drives the plane forward. This reaction force in the engines of our modern jet planes, as we have seen, may be as great as 10,000 horsepower. Note that the exhaust gases need not push against the air around the plane for this reaction force to exist, hence a jet plane can operate at altitudes where the air is much less dense than at sea level.

Since the interaction between two forces is sometimes very difficult to measure, as we saw in Fig. 5-4, we occasionally forget that there are at least two forces involved in *every* change in movement. The action force on a particular body is *always* balanced by a reaction force acting on another body. Since Newton himself experienced difficulty in applying this law to certain problems it is no wonder that this law may be confusing to some. However, in many simple situations the law can be applied easily.

With this study of Newton's third law we have now completed our investigation of a set of fundamental concepts of force. It is upon these three laws that a considerable portion of physical science is based. The first law tells us that all objects have inertia, or a resistance to being moved or changed in their movements. The second law shows how a force can be measured by observing its effect in accelerating a mass and how the force of gravity causes an acceleration of an object and thus gives it its weight. And, finally, the third law summarizes an important aspect of forces—that forces are interdependent and that no force can act alone.

5-6 Combinations of Force

As we have seen, force, since it has both magnitude and direction, is a vector quantity. Several forces acting upon an object, therefore, can be added according to the rules of vector addition. In Fig. 5-5 the two forces *A* and *B* are pull-

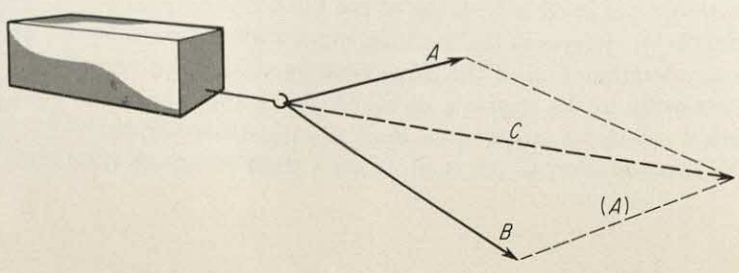


Fig. 5-5. The resultant of two forces acting upon an object is found through a simple vector diagram.

ing on the block. To find the resultant force, we put the tail of vector A, now (A), at the head of vector B, keeping vector (A) parallel to its line of force, and then draw a line C from the tail of vector B to the head of vector (A). The combined effect of forces A and B, then, is the same as if a single force C were pulling the block.

Example problem: In Fig. 5-6 a boat is being propelled directly forward with a 260-lb force by its motor. At the same time a strong wind is blowing diagonally across the boat. If the wind pushes the side of the boat with a 130-lb force, in what direction will the boat travel and what is the resultant force that moves it? In this experiment we will ignore the friction between the boat and the water.

Solution: Draw vectors F_m and F_w for the force of the motor and the force of the wind respectively, with their lengths proportional to the magnitude of their forces. By vector addition rules, find the resultant force F_r. In this vector diagram, 1 in. equals 160 lb; F_m, therefore, is $1\frac{1}{16}$ in. long and F_w is $\frac{13}{16}$ in. long. F_r should be $2\frac{1}{16}$ in. long, which means that the boat will move in the direction indicated with a resultant 340-lb force.

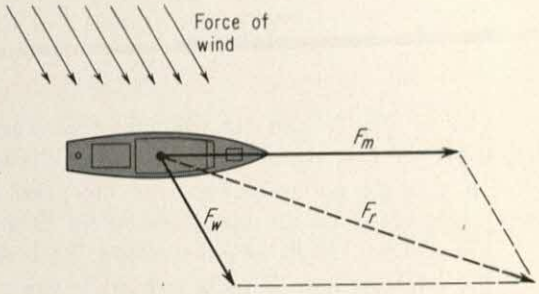


Fig. 5-6. The resultant force on a boat is a combination of the force of the motor and the force of the wind.

5-7 Forces in Equilibrium

When an object at rest is acted upon by several forces, it may either be set in motion or remain stationary. If the forces cancel each other out—if the resultant force is zero—the forces and object are said to be *in equilibrium*. The object will remain at rest if it is already stationary, or if it is moving, it will continue to move in the same straight line at constant speed according to Newton's first law. If, on the other hand, the forces from one direction are stronger than those from another direction—if, that is, the forces are *unbalanced*—the object will move in the direction of the resultant force.

In this case Newton's second law $(F = ma)$ comes into play, an acceleration will take place, and the object is not in equilibrium. So the first condition of equilibrium is that the sum of all the forces acting on a body must equal zero.

$$\Sigma F = 0 \qquad (5.6)$$

A tug-of-war is a good example of a balance of forces producing equilibrium. As long as the forces from one team equal those from the other, the rope stays still. But as soon as one team exerts more force than their opponents, there is some extra force which will produce an acceleration and equilibrium no longer exists.

To show how this principle helps us to solve problems, let us consider the case of a 160-lb man lying in a hammock as in Fig. 5-7. If the forces acting on the man do not equal zero, he will be accelerated—as in the case when one of the ropes break. When he is in equilibrium, however, we know that there is no resultant when we add all the forces together

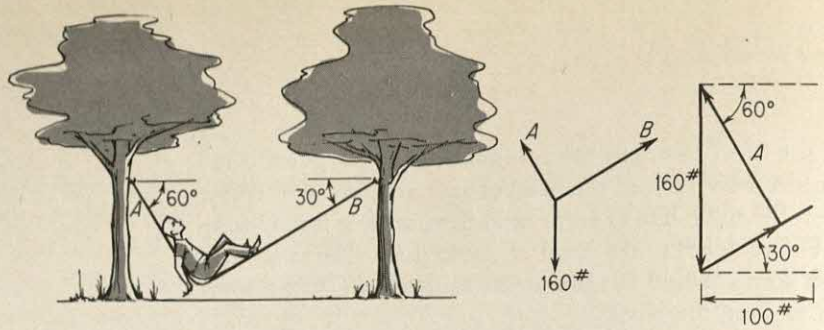


Fig. 5-7. Forces in equilibrium.

($\Sigma F = 0$). To find out what the forces are on each of the ropes we first sketch the direction of the forces as they act on a single point—in this case the point where all of the forces act is on the man (see vector diagram in Fig. 5-7). His weight, 160 lb, acts downward. Rope *A* pulls upward to the left and rope *B* pulls upward to the right. If we draw these three forces to scale as vectors there will be nothing left over to provide a resultant and our vectors will form a closed figure (a triangle in this case since there are just three forces). Since we know only the direction but not the amount of forces *A* and *B*, we will start with the weight of the man and draw a vector to scale, 160 lb straight down. To this vector we now add a vector representing force *B*. All we know, however, is the direction, not the length (or magnitude), so we leave the vector indefinite in length. Since we do know that these three forces form a closed figure ($\Sigma F = 0$), we know that force *A* must join vector *B* and the other end of the 160-lb weight vector. We therefore draw vector *A*, starting from the top of the 160-lb vector and proceeding backward in the proper direction until we intersect the line of force *B*. We now have a closed vector diagram, and we can measure these vectors to determine, with the scale we used, just how much force rope *A* and rope *B* must exert. If you measure the vectors using the scale in the figure, you will find that the rope *A* exerts a force of about 138.5 lb, while *B* pulls with about 80 lb.

Although we will not do much with it here, we ought to mention a second condition of equilibrium that is necessary to prevent a rotational acceleration. If a force is to produce rotation, it must act off center and the effect it has will depend upon the product of the force and its perpendicular distance to the point about which rotation might take place. We call this product a *torque* or *moment* (*M*).

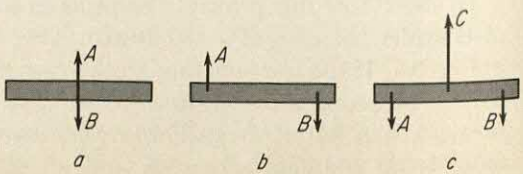
$$M = F \times l \qquad (5.7)$$

where l is the perpendicular distance from the line of action of the force to the axis of rotation. We can now write the second condition of equilibrium as

$$\Sigma M = 0 \qquad (5.8)$$

The examples in Fig. 5-8 may clarify this second condition of equilibrium. In Fig. 5-8a, the upward force *A* and the downward force *B* act at the same point, and being equal and opposite, produce equilibrium from the first condition

Fig. 5-8. Linear and rotational equilibrium.



alone; i.e., $\Sigma F = 0$. In Fig. 5-8b, we have force A equal to force B, but they are not acting at the same point and will produce a rotation, that is, they both produce a clockwise moment or torque about the axis and would provide a rotational acceleration; therefore, there is no equilibrium here. In Fig. 5-8c, the moments of A and B are equal and opposite in their effect upon rotational acceleration so the sum of the moments equals zero and the second condition of equilibrium is satisfied. To prevent linear acceleration, however, we can add a force C, acting at the center, equal to the sum of A and B so that the $\Sigma F = 0$. Fig. 5-8c will neither rotate or be accelerated linearly but is truly in equilibrium.

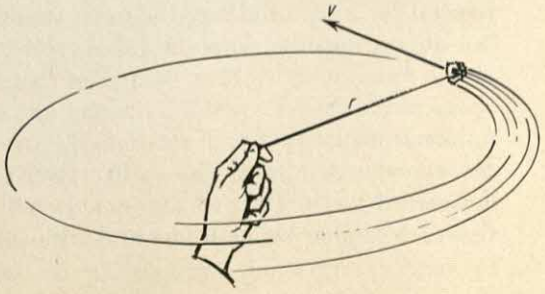
Notice that an object that is moving may be in equilibrium. For instance, an airplane traveling uniformly at 300 mi/hr has several forces acting upon it—the pull of the motor, the drag or friction with the air, the upward lift of the wings, and the downward force of gravity (weight) on the whole plane. If all these forces are in equilibrium, the plane continues to move forward at a constant velocity. A book lying on a table, a lamp hanging on a cord, and a car traveling at a constant velocity on a level highway are examples of bodies that are acted upon by forces in equilibrium.

5-8 Unbalanced Forces

When the forces that act upon an object are not in equilibrium, the object will be accelerated. For instance, a raindrop that is falling through the air is accelerated to higher and higher velocities by the force of gravity. A stone rolling down a hill, the water flowing over the brink of a waterfall, a car being accelerated to higher velocities, and a car being slowed down by the application of its brakes are all examples of objects that are being affected by unbalanced forces.

It often happens that an object is subjected first to unbalanced forces and then to forces in equilibrium. For example, as a raindrop falls downward faster and faster it encounters more and more resistance from the air, which acts as an upward force so that when the raindrop reaches a velocity of about 20 mi/hr the upward force of air resistance is equal to the downward accelerating force of gravity. The falling raindrop is then in equilibrium and it continues to fall at a constant velocity.

Fig. 5-9. Circular motion.



5-9 Circular Motion and Centripetal Force

We now have the tools to examine the complex motion of an object moving in a circle. A wheel rotating on its axle, a car rounding a curve in a road, and the earth rotating on its axis are all examples of rotational motion. If we whirl a rock on the end of a string, as shown in Fig. 5-9, we find that the rock's velocity vector v is at a right angle to the radius r of

the circle. A moment later, however, the velocity vector will have changed its direction, although its magnitude will have remained the same. We can conclude, therefore, that as an object travels uniformly in a circular path, its velocity will continually change, but its speed will remain constant. Even though the magnitude of a vector remains constant, if the direction changes, a force is required to cause this change of direction and an acceleration has occurred.

As the rock whirls around at the end of the string, what force causes it continuously to change its direction? If we let go of the string, the rock will continue in a straight line in the direction it was moving at the instant the string was cut. This is an example of Newton's first law. Therefore, to cause the rock to continually change direction we must exert a force on it toward the center at all times. This pull which we exert on the rock is called the *centripetal force.* But for every force there is an equal and opposite force and in this case the rock pulls on the string with this force which we call a *centrifugal force.* This force is given by the relationship:

$$F = \frac{mv^2}{r} \tag{5.9}$$

where r is the radius of the circle and v is the instantaneous velocity of the rock.

If you are a passenger in a car which is going around a curve, you seem to be thrown to the outside of the car and some people would say that centrifugal force pulled you out. This you now know is incorrect—it was Newton's first law in operation. You simply tried to continue on in a straight line. To cause you to go around the curve, the car exerted a force on you (centripetal force) and pulled you from your straight line path. At the same time you were pulling on the strap or pushing against the outside of the car with an equal and opposite force—the centrifugal force. Returning to the rock whirling on a string, Fig. 5-10 shows this action-reaction character of the two forces. The centrifugal force arises from the fact that the rock is constantly being accelerated toward the center of the circle by the inward force that we exert on the string; it is the reaction force to the centripetal force, and it arises from the tendency of the rock to move in a straight line.

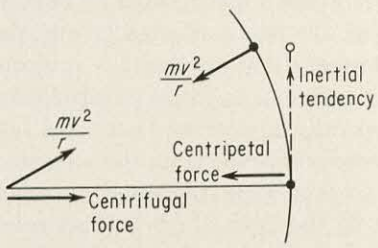


Fig. 5-10. The action-reaction forces in circular motion.

The bulge of the earth at the equator is caused by the centrifugal force that results from the rotation of the earth on its axis. The force of gravity provides the balancing centripetal force; if it did not, the earth would probably spread out almost into the shape of a disk. Newton was so certain of the validity of his laws of motion that he calculated the bulge of the earth's equator by taking into account the gravitational forces, the rate of rotation, the size of the earth, and the average density of the earth, which he guessed to be between 5 and 6 (it is almost exactly 5.5). His estimate of the earth's bulge is almost identical with the values measured by modern surveying methods. In the photograph of the

planet Jupiter (Fig. 7-6) you can actually distinguish the slight flattening which has occurred due to its high speed of rotation.

Since the force needed to keep a car going around a curve varies as the square of the velocity of the car, many drivers are fooled into thinking that a few miles an hour will make little difference. As you can see, it takes four times as much force to hold the car on the road around a curve at 60 mi/hr as it does at 30 mi/hr. Since $F = \frac{mv^2}{r}$, the force varies as the square of the velocity.

5-10 Kinetic and Potential Energy

Up to this point we have been considering forces that are associated with the mass, inertia, and the momentum of objects. These forces arise from two kinds of mechanical energy that are characteristic of all objects: kinetic and potential energy. Even though these two types of energy are closely related to each other, they represent two distinct aspects of mechanical energy that we will now examine.

Kinetic energy is the energy of motion. Any object that is moving possesses kinetic energy because it has the ability to do work, which, you will recall, is the capacity to exert a force through a distance. The man at a carnival who shows off his strength by striking an anvil with a mallet in order to send the bob up to ring the bell is trying to add enough kinetic energy to the mallet so that it will make the bob ring the bell. A baseball player swings his bat as fast as possible in order to impart as much kinetic energy to the ball as possible.

Common sense tells us that a locomotive traveling at 60 mi/hr has more kinetic energy than a golf ball traveling at the same velocity, and that a bullet has vastly greater energy when it leaves the muzzle of a gun than it does sitting in our hand. Thus, we can readily understand that kinetic energy is directly related to both the mass and to the velocity of an object. The mathematical equation that describes kinetic energy is:

$$K.E. = \tfrac{1}{2}mv^2 \qquad (5.10)$$

Note that momentum (described in Equation 5.2) and kinetic energy are similar in that both are directly related to mass and to velocity. Momentum, however, describes merely the motion of a particular mass, whereas kinetic energy, as shown in Equation 5.10 describes the capacity that a moving object has to do work. These are two different concepts. Let us trace the derivation of the equation for kinetic energy so that we can see why it is that the velocity must be squared.

When an object starts from rest and is uniformly accelerated through some distance s, we saw that the formula for

this distance was

$$s = \tfrac{1}{2}at^2 \qquad (3.7)$$

or solving for t^2:

$$t^2 = \frac{2s}{a}$$

Also in the case where the initial velocity is zero and the acceleration is uniform,

$$v = at \quad \text{or} \quad v^2 = a^2t^2 \qquad (3.3)$$

Again, solving for t^2:

$$t^2 = \frac{v^2}{a^2}$$

If we equate these two relationships

$$t^2 = \frac{2s}{a} = \frac{v^2}{a^2}$$

and solve for v^2,

$$v^2 = 2as \qquad (5.11)$$

Remember the special conditions: the object started from rest, was accelerated uniformly, and traveled a distance s during this acceleration.

Now let us consider the case of the work done in causing this object's motion. A force is needed to cause the acceleration ($F = ma$) and this force moves through a distance s. The work done then is:

$$\text{work} = Fs = mas$$

Since from Equation 5.11, $as = \tfrac{1}{2}\ v^2$, we can make the substitution and obtain:

$$Fs = \tfrac{1}{2}mv^2 \qquad (5.12)$$

This relationship shows that work (Fs) done on an object may result in motion and that the object in motion has an energy due to its motion ($\tfrac{1}{2}\ mv^2$). This we call kinetic energy. While in our example we did work (Fs) on an object, it is just as correct to think of the object doing work for us. For instance, a hammer strikes a nail changing its kinetic energy into a force which drives the nail. In stopping a car we must exert a force (by braking) through a distance and if the velocity of the car is doubled, the distance required for a particular force to stop it will be increased by four times.

Let us calculate the kinetic energy of a car with a mass of 1000 kg (weight about 2200 lb) going 30 m/sec (approximately 60 mi/hr).

$$\begin{aligned} K.E. = \tfrac{1}{2}mv^2 &= \tfrac{1}{2} \times 1000\,\text{kg} \times \left(30\,\frac{\text{m}}{\text{sec}}\right)^2 \\ &= 450{,}000\,\text{kg}\,\frac{\text{m}^2}{\text{sec}^2} \\ &= 450{,}000\,\text{joules} \end{aligned}$$

In contrast to kinetic energy, *potential energy* is the energy of position. A car at the top of a mountain possesses potential energy because it has the ability to do work by rolling down the mountain side. A clock spring that is wound up, the hammer of a pile driver lifted to the top of its frame, and water behind a dam all have potential energy. A hiker who has climbed to the top of a mountain has stored up considerable potential energy in his body by virtue of having lifted himself to the top of the mountain. Potential energy, like kinetic energy, is the capacity to do work and therefore is also measured in joules. The potential energy an object possesses is equal to the energy required to lift the object to its position. In Fig. 5-11 the mass m has been lifted from the floor to its position on the shelf. The work involved in lifting the mass is:

$$W = Fs$$

Since the force needed to lift a mass m is given by the equation:

$$F = mg$$

then:

$$W = mgs$$

The mass m, therefore, has mgs more energy on the shelf than it did on the floor, which means that its potential energy is:

$$P.E. = mgs \qquad (5.13)$$

where s is the height through which the mass has been lifted.

Using the illustration in Fig. 5-11, let us calculate the potential energy of the 10-kg mass which is 2 m above the floor.

$$P.E. = mgs = 10\,\text{kg} \times 9.8\,\frac{\text{m}}{\text{sec}^2} \times 2\,\text{m} = 196\,\text{joules}$$

In other words, if the 10-kg mass fell from the shelf to the floor, it could perform 196 joules of work upon striking the floor. As the mass falls, its potential energy is converted into kinetic energy. On the shelf it possessed 196 joules of potential energy; at the instant it strikes the floor, it will have 196 joules of kinetic energy. From this energy relationship we can determine how fast it will be going when it strikes the floor, as follows:

$$P.E. = K.E.$$

$$Fs = \tfrac{1}{2}mv^2$$

$$v^2 = 2\,\frac{Fs}{m} = 2\,\frac{mgs}{m} = 2gs$$

(Does this remind you of Equation 5.11? Notice too that Galileo was correct in saying that the mass of the object would have no effect on its rate of fall.) Solving the last equa-

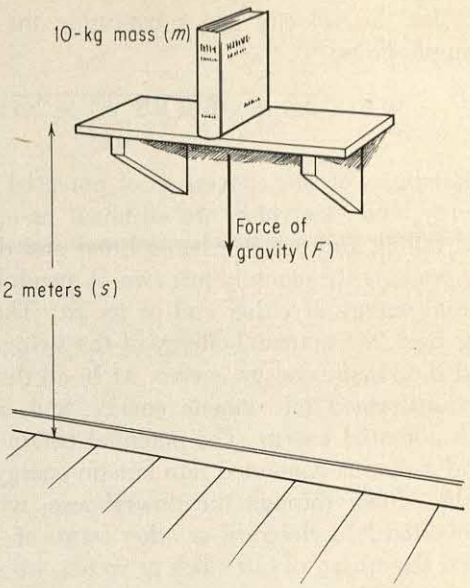


Fig. 5-11. The potential energy of an object is proportional to its mass and its height.

tion for the velocity and substituting the values from the example gives:

$$v = \sqrt{2gs} = \sqrt{2 \times 9.8 \frac{\text{m}}{\text{sec}^2} \times 2\text{m}} = 6.26 \frac{\text{m}}{\text{sec}}$$

Examples of this conversion of potential energy to kinetic energy, and vice versa, are all about us—an elevator rising and falling and our feet being lifted and dropped while we are walking, to mention just two. A pendulum possesses potential energy at either end of its arc. Thus, at point *A* in Fig. 5-12, the potential energy of the swing is at a maximum and the kinetic energy is zero. At *B*, all the potential energy is transformed into kinetic energy, and, at *C*, back again into potential energy. The potential energy of the water behind a dam is converted into kinetic energy when the water rushes down through the powerhouse, where it is usually converted into electrical or other forms of energy. When we wind the spring of our clock or watch, we store up potential energy in the spring, which, as the spring unwinds, is converted into the kinetic energy of the moving parts of the timekeeper.

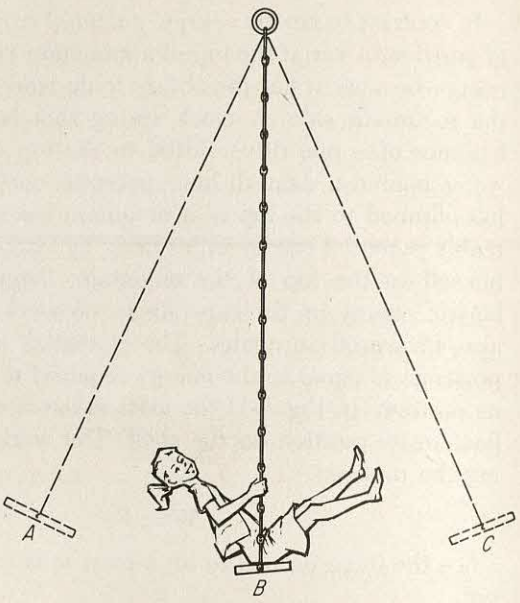


Fig. 5-12. The conversion of kinetic energy into potential energy (and vice versa) in a swinging pendulum.

5-11 Conservation of Energy

When one form of energy is changed into another form of energy, no energy is destroyed. This important principle—called the *conservation of energy*—has had far-reaching effects in the development of modern science. For instance, some of the energy of the swinging pendulum may be used up in friction with the air, but it is not destroyed. Instead it is converted into heat, which is another form of energy. As another example, the energy of a cue ball is imparted to other billiard balls when it strikes them. Although the movement of several balls may appear to involve more energy than the cue ball possessed originally, an analysis of the momenta of the balls before and after the collision would prove that no energy had been gained, or lost, in the process. In all the countless situations that have been studied by scientists involving the transfer of energy from one object to another and the conversion of one form of energy into another, no instance has ever been found where energy was gained or lost. From the activity on the minute scale of the nucleus of atoms to the grand behavior of the stars, scientists have found the law of conservation of energy to be valid. It is perhaps the most exact and the most firmly established law in all science.

Later on, when we discuss other forms of energy, we will have occasion to refer back to the principle of conservation. The breadth of this law is apparent when we follow the conversion of energy through its various forms from the potential energy of water behind a dam to the heat energy in a cup

of coffee. First, the water is permitted to fall from the dam through a turbine where its potential energy is converted into the kinetic energy of the whirling movement of the turbine. The turbine turns the generator and thus converts mechanical energy into electrical energy. The electrical energy is carried to our home through a powerline and, when passed into our stove, is converted into heat energy, which brings our pot of coffee to the boil. We can also convert the electrical energy into light, or even into chemical energy. The commercial production of aluminum is based on the conversion of electrical energy into chemical energy. The details of many of these energy conversions will be reserved for later chapters, but in all these conversions no energy is lost, although some of it is used in friction and appears as heat.

5-12 Summary

We have examined in this chapter the fundamental concepts of force as embodied in Newton's three laws of motion: (1) An object at rest tends to remain at rest, and a moving object tends to move in a straight line with constant velocity unless acted upon by an outside force, (2) forces are measured by observing their ability to accelerate a mass, (3) no force acts alone, and for any force there is always an equal and opposite force. This principle of action and reaction is basic to all processes wherever force is involved.

Inertia is the tendency of an object to continue doing what it already is doing. Momentum is the quantity of motion an object possesses; it is measured by the product of the mass and the velocity of the object.

When several forces act upon an object, their net effect may be found by adding the force vectors. A simple method of doing this is to construct a diagram in which the magnitude and direction of the forces are shown by suitable arrows. If the total effect of all the forces upon an object add up to zero, then we say that the forces are in equilibrium. If the forces are unbalanced, the object is acted upon by a resultant force that will probably cause the object to move.

If an object moves in a circular or orbital path, there is an inertial force that holds the object in such a path (centripetal force) and an opposite force produced by the tendency of the object to continue in a straight line (centrifugal force).

All ordinary energies of a mechanical nature are of one of two kinds: kinetic or potential energy. Kinetic energy is the energy of motion, and potential energy is the energy of position. Each of these may be converted into the other, or into any other form of energy such as electrical, heat, light, or chemical energy. In all conversions of energy from one form to another, no energy is ever created or destroyed. This law of conservation of energy governs all the processes that occur in nature.

EXERCISES

A. KEY TERMS TO REMEMBER

centrifugal force
centripetal force
equilibrium
foot-pound
force
horsepower
joule
inertia
kinetic energy
law of conservation of energy
momentum
Newton's laws of motion
potential energy
resultant force
unbalanced force
weight

B. QUESTIONS ABOUT FORCE AND MOTION

1. Is the inertia of a train moving at 60 mi/hr different from its inertia when it is standing in a station?

2. Explain how the moon, in moving in an orbit around the earth, is constantly being accelerated toward the earth.

3. In hiring a laborer, would you rather pay him on the basis of his power or on his total work?

4. Describe in general how you could determine the mass of an object by using Newton's second law if you were located out in space beyond the effects of gravity. How may this law be used to determine the mass of an object on the surface of the earth?

5. If the speed of the earth's rotation were somehow increased, what would happen to the weight of an object (a) at the equator? (b) at the pole?

6. A man, caught at the center of a smooth, perfectly frictionless ice pond and without any outside help, extricated himself by throwing his shoes southward, whereupon he started moving slowly northward. Is this method of escape plausible, and which of Newton's laws apply?

7. Indicate which form of mechanical energy, kinetic or potential, is primarily involved in each of the following:

a. a rolling log on a horizontal surface
b. a log rolling down a hill
c. a tightly-wound watch spring
d. a stretched rubber band
e. a spinning fly wheel on an engine
f. a cocked mousetrap

8. By a diagram show the conversion of one form of energy into another when a stone rolls down a hill and plunges into a lake at the bottom. Try to account for all energy changes, keeping in mind the law of conservation of energy.

9. A rocket puts a satellite into orbit from Cape Kennedy so that it passes over the Atlantic and Africa. Would this affect the rate of the earth's rotation? Would it make any difference if the satellite had been shot from a gun?

10. If every action has a reaction, a wagon must pull back on a horse just as hard as the horse pulls ahead on the wagon. Under this circumstance, how can the wagon ever move?

11. By means of a diagram, show the acting and reacting forces involved when a person sits in a chair.

12. A car traveling due north is increasing its rate of travel continuously. The resistance of the car to this increasing rate of travel is described *best* by using the term (a) acceleration, (b) speed, (c) velocity, (d) inertia.

13. Acceleration is (a) inversely proportional to force, (b) directly proportional to force, (c) directly proportional to mass, (d) equal to force times mass.

14. "Every force that acts upon a body always has an unequal and opposite force opposing it." Which word or phrase is *incorrect* in this statement of Newton's third law? (a) every force, (b) always, (c) unequal, (d) opposite.

C. PROBLEMS ABOUT FORCE AND MOTION

1. Calculate the horsepower needed to lift a 10-lb ball to the top of the Washington Monument, 555 ft high, in 5 sec. Ans. 2.02 horsepower.

2. Make a diagram of a bowling ball at the instant it strikes a bowling pin. Show by simple vectors the resulting forces.

3. Show by vectors all the forces involved when two boys, holding opposite ends of a rope, pull on it with equal force.

4. By using vectors, find the direction and amount of force acting upon a stone that is being pushed by two boys, one pushing northward with a force of 50 lb and the other pushing eastward with a force of 25 lb. Ans. 55.6 lb of force north of northeast.

5. A 5-kg mass at rest is to be accelerated at 2 m/sec^2 until it gains a velocity of 20 m/sec.

a. What force is necessary to give this mass its acceleration?
b. How long a time will it take for this force to accelerate this mass to 20 m/sec?
c. How far will the mass move during this acceleration?
d. How much work does the force do?
e. What kinetic energy does mass have after acceleration?

6. How much more work is necessary to stop a car going 60 mi/hr than to stop it from 20 mi/hr? from 30 mi/hr?

7. A car rolls over a cliff 121 ft high and crashes on the rocks below. At its instant of impact, it has the same kinetic energy as if it were going on a level road at what speed?

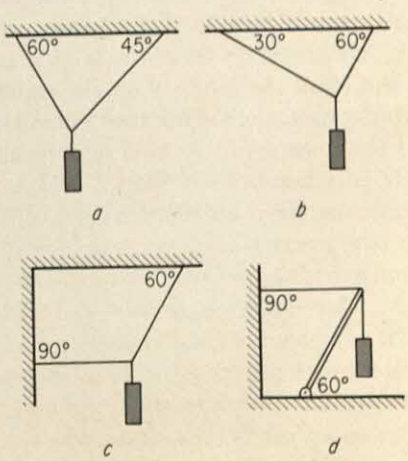
8. A bullet weighing 0.20 kg leaves the barrel of a gun 0.80 m long with a velocity of 400 m/sec. If its acceleration was uniform, what average force was exerted on the bullet? Ans. 20,000 newtons.

9. A man weighing 80 kg stands in an elevator. What force does he exert against the floor of the elevator when it:

a. is stationary?
b. is accelerating upward at 2 m/sec^2
c. is moving at a constant velocity of 20 m/sec upward?
d. is falling freely (if the cable breaks)?

10. Draw the force diagram showing the tension in each cord in Fig. 5-13 if the weight in each case is 100 lb.

Fig. 5-13.



Six

Force and Motion in Natural Processes

In Chapter 4 we saw how the force of gravity, acting upon streams and glaciers as they move down an inclined plane, change the earth's surface. In the last chapter we examined the fundamental laws of motion and learned about several important relationships between force and motion. In this chapter we will consider the movements that take place beneath the earth's surface, movements that profoundly affect the surface appearance of the earth. We will find ample proof that many parts of the earth's crust have been displaced and distorted in movements that require tremendous forces. We shall examine some of the evidence of these movements, and try to find the source of the forces that cause these movements.

A piece of coal may carry a fossil imprint of a leaf, a twig, or even a whole tree. Yet it may have been found in a mine several hundred feet deep. A highway engineer, cutting into a rocky cliff high up on the side of a mountain, may come upon the fossil markings of snails, clams, and other marine life which indicate that the rock containing them was at one time below sea level. A motorist traveling through a hilly country sees hundreds of feet of rock layers that were once sediments at the bottom of a sea. What enormous forces pushed the mountains up from the ocean floor? How did the coal, once made up of plants growing in the sun, come to be buried under hundreds of feet of rock? Before we investigate the possible answers to these questions, let us examine the structure of the earth's crust as it now is in order to more fully understand the forces and motions that have produced the mountains and the lowlands.

6-1 Warps

In Chapter 4 we saw how sedimentary rocks, such as limestone, sandstone, and shale, are always formed at the bottom of a lake or sea. The force of gravity causes the sediments to settle out and form a horizontal layer of material. Subsequent loads are deposited on top of these sediments and compact them into sedimentary rocks. So reasonable and dependable is this idea that geologists for over two centuries have agreed that it is a fundamental principle of nature. This concept, called the *law of original horizontality*, is based upon the premise that all strata of sedimentary rocks were originally formed in a horizontal, or very nearly horizontal, position. Consequently, when we find a layer of sedimentary rock that is no longer in a horizontal position, we must conclude that it has been disturbed. The next time you pass through a road cut and see the rock strata, by a brief examination of their positions you will be able to read at least a part of the story of their past. If they are horizontal, they have probably not been disturbed since they were formed, but if they are tilted, wrinkled, or distorted in any way, they have been disturbed since they were formed.

The gentlest disturbance of rock layers is called a *warp*, which is a rock strata that is slightly off the horizontal. The Central Plains of the United States are, for the most part, composed of rock strata overlain with soil. Although these strata often appear to be horizontal, careful measurements show that many of them slant slightly upward or downward. For example, in eastern Kansas the limestone strata slope downward to the west at the gentle rate of 3 feet a mile, and it is interesting to note that the present topography of this region slopes downward to the east.

If we were to get a huge cross section from east to west through the Central Plains by slicing downward several thousand feet, we would discover many broad gentle warps. From such a slice, say, from Manitowoc, Wisconsin, on the west shore of Lake Michigan straight east to Kitchener, Ontario (Fig. 6-1), we would see that even though the surface of this country is fairly level, underneath it the original sedimentary rocks are now downwarped. Beginning at the surface at Manitowoc we may follow the old sedimentary rock downward to the east until almost under Midland, Michigan,

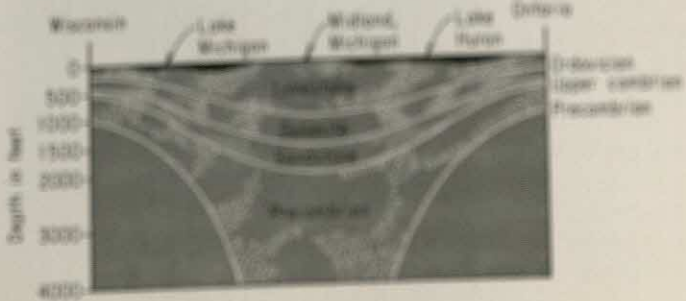


Fig. 6-1. A cross section of the rock layers under the state of Michigan comprising the Michigan basin. (Drawn from information after Colton, *Bulletin of the American Association of Petroleum Geologists*, 32 (1948), 1417.)

we come to the lowest level of this old rock layer, now about 13,000 feet underground. If we could examine a similar slice made in the north-south direction through Midland, we would see almost the same general downwarp. That is, underneath the present level country there is a layer of rock that forms a shallow bowl, called the Michigan *basin*. This basin was formed in the Devonian period, sometime around 500 million years ago, and at that time similar basins were also formed in the Indiana-Illinois-Kentucky area, in the West Virginia-Pennsylvania area, and in the Nevada-Idaho area.

After the Michigan basin started to form, it provided a natural drainage area into which flowed streams laden with erosional products. In certain parts of this basin, particularly the deeper part near Midland, huge deposits of salts were formed. Some of these salt beds are over 2000 feet thick and now provide the raw materials for many industries located in this area.

In contrast to basins produced by downwarping are concave uplifts called *domes*. In the central part of Tennessee is the Nashville dome, in which the subterranean rock layers of the Devonian period are near the surface, sloping gently downward as one travels in any direction away from Nashville. The gentleness of the warps in this area are indicated by the gradual slope of these rock layers—8 to 16 feet a mile.

Advancing and receding coast lines indicate that parts of the earth's surface are rising and falling even today. Warping, then, is a process that goes on all the time. The eastern coast line of the United States is advancing landward and is thus marked by bays and irregular estuaries. The bottoms of these bays actually contain sediments derived from dry-land plants and animals—proof that the bays were once above sea level. In other parts of the world, on the other hand, coastal areas are rising, causing the coast line to advance seaward. We know the coasts of Alaska and Scandinavia have risen because the exposed rock along these coasts contains marine fossils.

The explanation for the upwarping and downwarping of various parts of the earth's surface is not fully understood. In some instances, the movement appears to be merely a matter of re-establishing isostatic balance. The Scandinavian coast line is emerging supposedly because it was weighted down with a huge glacier during the age of continental glaciers. It has been estimated that since the last glacier melted away eleven thousand years ago the coast lines of Sweden and Finland have risen 900 feet. If the glacier that is today on Greenland were to melt, we would expect the removal of the 8000 feet of ice to reduce the downward force on the land and to permit it to rise. With the melting of all this ice the level of the oceans would be raised some 20 feet, which in turn would add to the weight of the oceans and set up new forces in the earth's crust that would require new warps in order to return to equilibrium.

Upward and downward movements of noticeable extent

have occurred within the time of recorded history. The temple of Jupiter Serapis at Naples, Italy, was originally built on land near the coast. All that remains of this ancient building are three columns that now stand partially submerged in the sea, but the lower 18 feet of these columns, above the present waterline, are covered with marine deposits. Apparently the coast line there sunk and then rose, all within about two thousand years.

A revealing test of the isostasy principle was made when Hoover Dam was built. The water behind the dam forming Lake Mead is a downward force on the earth's crust that has been added in a relatively short time and it has been found to have caused the rock on either side of the lake to be elevated 11 feet, thus producing a man-made warp in the earth's surface.

6-2 Folds

Ordinarily we think of rock layers as being quite rigid, whether stressed in a vertical or in a horizontal direction. Rock such as limestone and sandstone is specially quarried and used in buildings because it is capable of bearing heavy loads. But within the earth's crust, rock exhibits a plastic or fluid character if subjected to great forces. This is evident from observations of rock layers that have been compressed horizontally into what are called *folds*. These folds, or wrinkles, may be less than a fraction of an inch wide or they may be several miles in size. If we push the edges of a tablecloth together the cloth becomes compressed into many small folds. Similarly, the earth's crust in many places appears to have been compressed latterly into upward and downward folds; one of these that has been exposed is shown in Fig. 6-2.

The upward arch of a fold is called an *anticline*, and the downward part, or the trough, is called a *syncline*. Obviously, the anticline is the more accessible part of the fold for ob-

Fig. 6-2. A single upward fold in the earth's crust. (U. S. Geological Survey)

servation because the syncline is buried beneath the surface. The anticline, being forced upward, is subject to erosion, and often the rock in the anticline is so disrupted, cracked, and distorted in the folding process that it may be readily eroded away. When this happens, the upper parts of the adjacent synclines may be exposed, and the structure of a region can be more carefully studied. Oil well drilling records, highway cuts, and stream erosion provide other sources of information through which a fairly satisfactory idea of the arrangement of the rock layers in synclines can be obtained.

Folding in the earth's crust accounts partly for the formation of the Appalachian Mountains, the Swiss Alps, and the Himalayas. In the Appalachians, the folds were originally formed during the Appalachian Revolution, roughly 200 million years ago. These folds, running in a northeasterly-southwesterly direction cover a good-sized portion of the eastern part of the United States between Alabama and Georgia on the south and New England and Nova Scotia on the north. In traveling east to west through these mountains, we can readily see the remnants of fold after fold that have been formed in the earth's crust. For instance, if we drive through Pennsylvania on the Turnpike, we pass through five tunnels, each almost a mile long, that burrow through the series of folds that make up these mountains. If all the folds in the Appalachians were smoothed out, the distance we would need to travel going east to west would be increased about 40 miles. West Virginia hill folk jokingly claim that if their state were ironed out flat it would be as big as Texas.

The present surface of many parts of the Appalachians no longer corresponds to the original folds. In recent geological periods the folds have become stabilized, that is, they have not been distorted, pushed together, or pulled apart significantly. Erosion has removed most of the ridges and filled in the troughs to the extent that many of the present ridges are formed by erosion-resistant rock that at one time lay in the lower part of a fold. An examination of Fig. 6-3, which shows the present surface and subsurface structure of Limestone Ridge in Perry County, Pennsylvania, shows how erosion has greatly changed the landscape. The top of the original fold is now eroded away and the actual surface contours are now determined, not so much by the shape of the original fold, but by the different rates of erosion of the various parts of the fold.

The Alps of Switzerland are partly formed of folds, but the folds are more complex than those of the Appalachians. In the Alps, some areas have been compressed laterally so

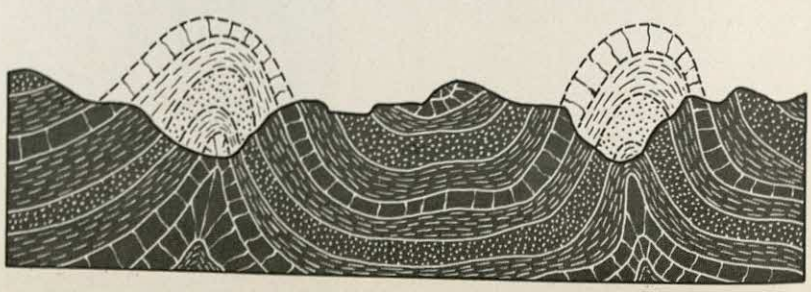


Fig. 6-3. Cross-section of an Appalachian fold in Pennsylvania. (From a geologic map of Pennsylvania, Pennsylvania Topographic and Geologic Survey, 1931)

much that the folds have overturned, causing certain parts of the fold to contain rock layers that are upside down compared to their original position. Before the concept of folding was understood, these upside-down rock layers presented a difficult puzzle for the early geologist. The folded structure of the Alps has since been studied extensively, and quite satisfactory structures have been worked out, even for the most complex folds. The rocks that are folded in the Alps and the Himalayas are of Cenozoic Age, showing that these folded mountains are much younger (60 million years old) than the Appalachians. It is possible that the Appalachians were at one time as high or higher than the Alps, but because of their age they are now worn down by erosion to a fraction of their original height.

6-3 Geosynclines

A broad syncline several miles wide is often overlain by a broad flat valley that has accumulated a large amount of sediments. The whole structure—the downward warped rock layers and the accumulated sediments—is called a *geosyncline*. A good example of a geosyncline is the Central Valley of California, which is almost 100 miles wide and contains 25,000 feet of sediments. Although this valley has sunk as it has collected sediments, the rate of subsidence has been nearly equal to the rate of accumulation of sediments, so the valley has remained about the same distance above sea level —an excellent example of the isostatic principle. Geosynclines are evident in Nevada, western Utah, Idaho, and eastern Oregon, as well as in various other parts of the world.

In the geological past, there have been large geosynclines in many of the areas that are now mountainous. For instance, the Appalachian Mountains were formed out of the sedimentary rock layers that were originally desposited in a huge geosyncline during the early Paleozoic era, four or five hundred million years ago. At about the same time, an even larger geosyncline was building up sedimentary rocks in the region now occupied by mountains in the western part of the United States. Almost all major mountain ranges have been constructed of geosynclinal rock layers, and it is generally believed that this is more than coincidence, although no satisfactory explanation has been found for the fact that many mountains rise up from geosynclines. Several theories have been suggested for the process by which mountains arise from lower and relatively flat country, and we shall consider these in a later section of the chapter.

6-4 Faults and Joints

Cross sections of the earth's crust usually reveal a much more complex structure than that caused just by warps and

folds. When the forces in the earth's crust are too great or the rocks are too rigid to cause folding, then abrupt distortions occur. Such a rupture across the layers of rock is called a *fault*. Faulting is more complicated than folding in that it may be the result of either compressional or extensional forces, and when conditions are such as to produce folds and faults, very complex structures are formed.

There are primarily three types of faults, as shown in Fig. 6-4. A *normal fault* is the result of a rupture in a rock formation where the earth's crust is under tension. Forces tend to pull the rock layers apart, and in order to move, the rocks form a line of departure that is tilted. On one side of the fault line the rock moves downward, and on the other side the rocks move upward.

A *thrust fault* is the result of compressional forces in the earth's crust. These forces are probably similar to those causing folds, but when the rock is too brittle to fold, it makes an abrupt break. One side of the fault moves upward and over the other side, which is forced downward and under the first side.

In a *strike-slip fault*, the rocks on both sides of the fault line remain nearly at the same level but one moves, say, east while the other moves west. The forces that cause a strike-slip fault are probably the result of distortional stresses that build up in the earth's crust in the form of potential energy, which finally reaches a point where the rocks are no longer strong enough to preserve the stress. A grinding, grating slippage then occurs that may cause the earth in the vicinity to tremble mightily. What happens is that the potential energy is suddenly converted into kinetic energy.

The tension in the earth's crust is often relieved by the rock separating, but without the segments moving sideways with respect to each other. When rock structures are pulled directly apart in this manner the opening between the two sides is called a *joint*. A joint may range in width from only a fraction of an inch to several feet. It usually fills up with sediments, or, in deep underground processes, it may become filled with minerals of various types that are called veins. Almost all exposed rock exhibits jointing and faulting to some extent. Figure 6-5 shows the intersection of two fault lines in a series of limestone strata on the banks of a small stream.

Much of our most scenic mountain landscape is the result of faulting and jointing. Although the mountains in the west are very complex, part of their formation can be traced to thrust faulting of the Cordilleran geosyncline about sixty million years ago. In contrast to the old Appalachians, which are folded mountains, the beautiful mountains of Glacier National Park are probably the remains of a massive thrust fault known as the Lewis overthrust. Many of the Sierra Nevadas and the mountains of the Nevada-Utah area have not been seriously eroded, so that we can still identify the structure of the faults that produced them.

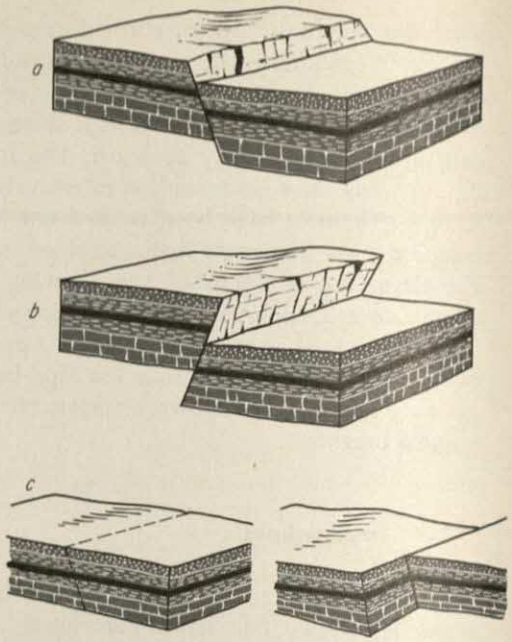


Fig. 6-4. Types of faults: (a) normal, (b) thrust, and (c) strike-slip.

6-5 Earthquakes

To observe faults and joints that have been formed in the dim past in order to understand the structure of the earth's rocky crust is a relatively quiet activity, but it is quite another thing actually to be near a fault when the rocks snap apart and produce an earthquake. Although we hear about only a few of them, each year there are over a million earthquakes on our planet, of which probably a dozen are strong ones. The earth, however, is shaken violently only about 200 times each century. The most severe earthquake in recorded history occurred in 1755 just off the coast of Lisbon, Portugal. The shock demolished most of the city and damaged buildings in cities within a radius of several hundred miles. A 50-foot wave, roiled up in the ocean by the initial shock, rushed into Lisbon, and, along with the original mechanical shock, was responsible for the loss of one-fourth of the city's population. The earthquake was attributed to the sudden slippage of a submarine fault somewhere out in the Atlantic Ocean.

In April, 1906, the famous San Francisco earthquake caused severe damage to that city and to others in the vicinity, but at the same time it considerably advanced our understanding of earthquakes. Prior to this earthquake an old strike-slip fault, called the San Andreas rift, had been studied carefully. This fault zone, several hundred miles long, had been a region of periodic activity, and it was surveyed with the object of finding out about the changes in the earth's surface caused by earthquakes. The 1906 rupture was again a strike-slip fault that occurred along part of the San Andreas rift, and in some places there was up to 21 feet of lateral displacement. Property worth nearly 500 million dollars was destroyed or damaged, and over seven hundred lives were lost.

Fig. 6-5. A geologist measures the inclination of two fault lines exposed in the rocky bank of a small stream. (Humble Oil & Refining Co.)

The fault that is the source of an earthquake usually extends down into the earth's crust some distance, for the center, or focus, of an earthquake is seldom near the surface, and it has been detected as deep as 400 miles under the surface. The point on the surface directly above the focus is called the *epicenter*, and it is this point that suffers the greatest shock. The energy associated with faulting is enormous. In the San Francisco earthquake, an estimated 10^{19} joules or 7.38×10^{18} ft-lb of work (force times distance) was performed by the potential energy stored up as tension and stress in the rocks on either side of the fault, and this work is equivalent to one thousand super H-bombs. This energy is not concentrated at one place but is released over many miles of the fault and deep within the earth's crust. Earthquakes probably produce as much energy in as short a time as any of the natural processes on our planet. It has been estimated that during the last hundred million years, the San Andreas fault has accumulated a lateral displacement of 300 km.

6-6 The Earth's Structure from Earthquake Studies

Although earthquakes cause great damage, they have, ironically enough, given us invaluable information about the structure of the earth. The science of earthquakes is called *seismology*, and the instrument that is designed to measure the intensity, direction, speed, and shock effects of an earthquake is called a *seismograph*. This device is essentially a pendulum suspended from a rigid support. Any tremor in the earth moves the support but the pendulum, because of its inertia, remains stationary. A stylus on the pendulum records the tremors in the earth's crust so that a permanent record of them is continually being made.

Earthquake shocks travel through the various parts of the earth by wave motion. In Chapter 10 the subject of wave motion is described, but for our present purpose it is enough to know that two kinds of mechanical waves travel through solid material—primary (longitudinal) waves (P) and secondary (transverse) waves (S). P-waves travel about 5 mi/sec, and S-waves travel about 60 per cent of the speed of P-waves. When these two types of waves are detected at a seismographic station, the lag of the S-waves behind the P-

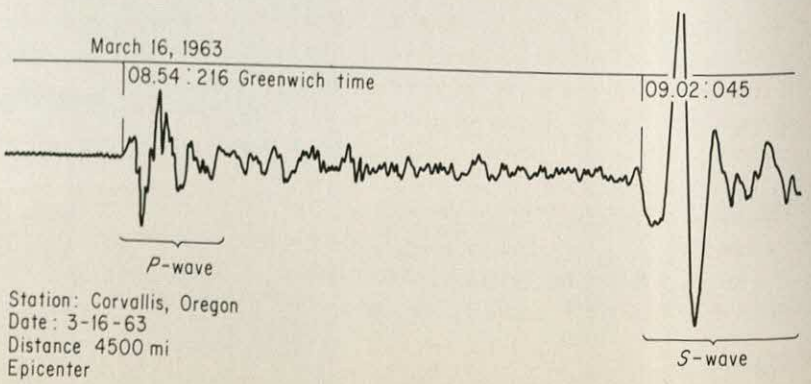


Fig. 6-6. A seismogram made at Corvallis, Oregon on March 16, 1963 produced by an earthquake of magnitude 7, the center of which was in the Kurile Island region, some 4,500 miles away. (Courtesy of Joseph Berg, Oregon State University)

waves is a clue to how far it is from the station to the point of origin (epicenter). Fig. 6-6 shows a difference of about 8 minutes between the arrival time of the P- and S-waves, and the distance to the epicenter can be estimated. Along with similar data from other stations, the position of the earthquake can be located.

In addition to the P- and S-waves in the earth, there are two kinds of surface waves that travel at slower rates and consequently arrive at a seismograph after the P- and S-waves. The surface waves continue to travel around the earth, and the Chilean earthquake in May, 1960, was so severe that surface waves were detected several days after the quake. The earth even vibrated at 20-minute intervals for more than a month.

From earthquake studies, the structure of the interior of the earth has been found to consist of three zones (Fig. 6-7 and Table 6-1). The outermost, and thinnest, zone is the *crust*, which is about 25 miles thick. Most of the dry land areas of the world are covered with a mile or two of sedimentary rock, and the rest of the crust under these areas is granite and basalt, two rocks that are formed as the result of high temperatures. There is probably little or no granite under the oceans—only basalt. Of course there are places where granite and basalt are exposed on the earth's surface, but in general they are covered with sedimentary rock.

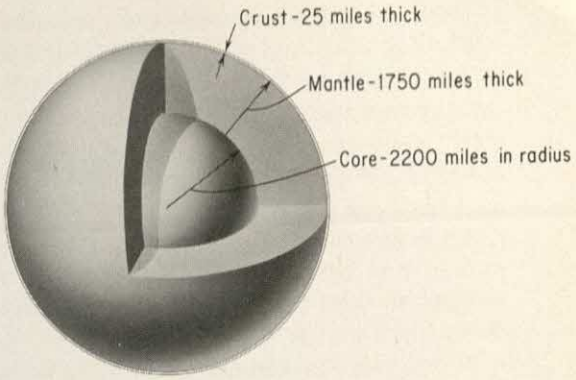


Fig. 6-7. Model structure of the earth based on earthquake studies.

TABLE 6-1 SOME CHARACTERISTICS OF THE INTERIOR OF THE EARTH

Earth's layers	*Depth (miles)*	*Velocity of P-wave mi/sec*
Crust	0–25	variable
Upper Mantle	25–600	5–5.5
Mantle	600–1700	5.5–8
Lower Mantle	1700–1800	uncertain
Outer Core	1800–3100	5–6.5
Transition Region	3100–3200	uncertain
Inner Core	3200–4000	7

Under the crust is the *mantle*, which is about 1750 miles thick. Not too much is known about the composition of this zone except that it is probably uniform in composition and has a temperature of about 3000°C. Its upper part may be glassy or almost a fluid, although not in the ordinary sense of a liquid that flows easily. In addition to high temperature, the material in this zone is subjected to extremely high pressures (possibly 10,000 pounds per square inch at the bottom of the mantle). Since we do not know much about the behavior of materials at these extreme conditions, it is difficult to describe them in terms of everyday experience.

The innermost zone of the earth, the *core*, is about 2200 miles in radius. It is made up of much more dense material

than the rest of the earth—scientists now generally agree that it is mostly iron—and its average temperature is estimated to be 5000°C. Since S-waves do not pass through the earth's core, at least the outer part of it must have liquid characteristics (S-waves are not capable of passing through a liquid), but the inner part is probably solid. The old idea of a liquid core has been repudiated by observations of the earth's momentum of rotation. That is, if the core, which is the densest part of the earth, were a liquid its lack of rigidity would permit the tidal friction caused by the moon to slow down the earth's rotation at a faster rate than it actually does.

Fortunately, in addition to earthquake waves, there are other means of studying the structure of the earth's interior —its momentum of rotation, its temperature, its composition and density, and its place as a planet in the solar system and the universe. As we encounter further information and methods for studying the universe in later chapters, we will again consider the problem of the earth's structure. Of one thing we are certain, there is much we do not know about the earth.

6-7 Metamorphism

At one time or another, many of us have placed a penny on a railroad track, stood back, and watched a train pass over the coin. Although the composition of the penny is not changed, its shape, size, and surface markings become greatly altered. Instead of a circular disk, the coin is pressed into a thin, elliptical object that has only a trace of its original markings. The penny, in short, has undergone a process of *metamorphism*, which means its characteristics have been changed by the application of great pressure, Fig. 6-8 (middle). In addition to pressure, metamorphism may also involve the effects of heat and solution, but they will be discussed in later chapters. Since metamorphic processes occur well down in the earth's crust, we can never examine metamorphism in action, but as in the case of the penny above, we can examine rocks both before and after they have been subjected to it. Sedimentary rocks on the surface are often forced downward by folding or faulting, and the great pressure at the lower depths causes distortions and changes in the structure of the rocks, as shown in Fig. 6-8. The crystals embedded in the rock may become elongated and thin, or they may become twisted and fractured, depending upon the severity of the metamorphic process.

In Chapter 4 we considered some of the common sedimentary rocks—shale, sandstone, limestone, and coal. When these parent rocks are subjected to high pressures and temperatures, they become metamorphosed into daughter rocks that have the same or similar composition but a considerably different physical appearance. Table 6-2 gives the relationship between parent and daughter rocks.

Fig. 6-8. Forces in the metamorphic processes within the earth's crust.

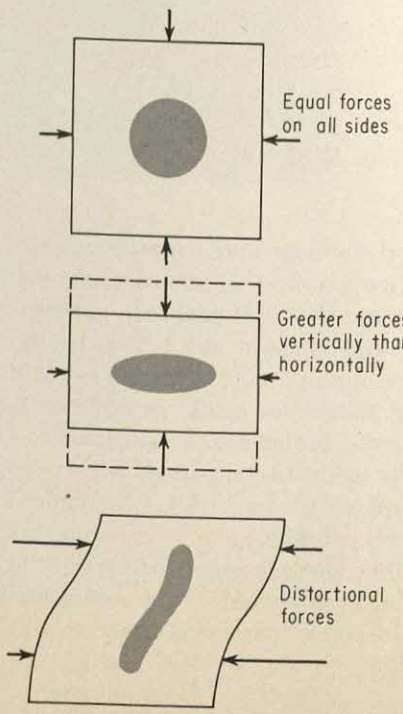


TABLE 6-2 **RELATIONSHIPS BETWEEN SEDIMENTARY AND AMORPHOUS ROCKS**

Sedimentary Rocks		
Fragmentary	*Precipitated*	*Organic*
Shale	Limestone	Bituminous coal
Sandstone	Chert (silica)	

Metamorphic Rocks	
Foliated	*Unfoliated*
Slate (from shale)	Marble (from limestone)
Schist (from shale)	Quartzite (from sandstone and chert)
	Anthracite (from bituminous coal)

A foliated rock, such as *slate*, when shattered with a sharp blow breaks into layers which indicates that it was formed primarily by a direct, downward pressure, as in Fig. 6-8 (middle). The original laminar appearance of shale is still apparent in slate, but the layers are much more compacted. Further metamorphosis of slate produces a *schist*, a rock in which there is no trace of the original structure of the shale, except that it is made up of extremely fine layers. In contrast to foliated rocks, unfoliated rocks show no layered structure. When shattered they break haphazardly across crystals and follow no particular pattern of rupture. *Marble* is limestone that has been metamorphosed by heat and pressure, and it often shows variegated or veined coloring that hints of the process in Fig. 6-8 (bottom). *Quartzite*, formed by both heat and pressure from either sandstone or chert, is a homogeneous rock much more compact and harder than its parent rock. In *anthracite*, the original sedimentary lines are still evident, but the structure has been greatly compacted.

6-8 The Origin of Mountains

Mountains, like people, are often reluctant to reveal their origins. Usually they have been formed by complex processes involving all kinds of displacements—warps, folds, and faults—in enormous crustal movements that would make the 1906 earthquake seem puny indeed. Following these crustal movements, erosion by water and ice further obscure the structure of the original mountain. The geologist, consequently, is usually very cautious in describing the process by which a particular mountain was formed. Let us now consider the major processes of mountain formation and examine a few examples of each.

Erosional mountains. At first, it would appear impossible to produce a mountain by erosion, but if we recall what a mountain is—a region that is noticeably elevated above its surroundings—we soon see how erosion can make a mountain.

If a large area, such as the Colorado Plateau, the Ozark Plateau in Arkansas, or the Catskill region in New York, is upwarped, subsequent erosion may remove part of these flat, upland regions and leave uneroded areas of mountainous size.

The Catskill Mountains were formed by erosion. The original plateau was 5000 feet above sea level, and the bottoms of the valleys are about 2000 feet above sea level. These 3000-foot mountains are composed of nearly flat layers of sedimentary rocks, and since the end of the Paleozoic time, erosion has been the primary change in this region. All the mountains have sedimentary layers that correspond to each other, showing that they were all part of the same land mass at one time. The Ozark Plateau, particularly in its western parts, is composed of erosional mountains, and here again the sedimentary rocks have not been disturbed except through the gradual process of erosion. Figure 6-9 shows a group of horizontal layers of sediments that has been eroded into a mountain.

Accumulation mountains. The most common mountain produced by accumulation is a volcano. A volcano is produced by the heat within the earth's crust, and since we can more fully appreciate its formation after a study of heat, we will reserve a more detailed consideration of volcanoes until a later chapter. For our present purpose, we can say briefly that a volcanic mountain is a deposit of rock, lava, and dust that has been spewed out of the volcano.

Diastrophism mountains. Diastrophism is the word we apply to any or all of the mechanical changes in the earth's crust—folding, warping, jointing, and faulting. Naturally, mountains formed by diastrophic processes can be very complex, because upwarping can be followed by folding, followed by faulting and more folding and jointing, and so on. The Appalachians are good examples of folded mountains, as we have already seen. Yet, in this range of mountains, which is

Fig. 6-9. Scotts Bluff near Gering, Nebraska, a mountain produced primarily by erosion. (Union Pacific Railroad photo)

primarily a group of folds, there is much faulting, particularly in the southern parts. For example, in the sharp folds of eastern Tennessee, nine major thrust faults have been detected and measured, all within a distance of only 20 miles. Most of these faults have the overlapping wall facing the west, indicating that the force which caused these faults came from the east and pushed the upper part of the earth's crust to the west. There was probably a succession of three main events in this region—folding, followed by faulting, and then finally more folding. The fault lines themselves are twisted, showing that folding followed the faulting, and the rock layers show two groups of folds, one formed after the fault and the other before.

In the basin ranges in Nevada, Arizona, and Utah there are many fine examples of mountains produced primarily by faulting. The Sierra Nevadas, which are among the highest mountains in the world (those bordering Death Valley are over 11,000 feet above the floor of the valley 10 miles away) are mostly the result of a fault tilted downward toward the west. That is, the region to the west of the fault line (or possibly several fault lines) has moved upward producing mountains that are over 14,000 feet above the region east of the fault line. A few mountains in this region were formed by thrust faults, the overlapping face of the upper side still being exposed as evidence of the type of faulting that produced the mountain. Most of the faulting, however, is the result of tensional forces in the earth's crustal layers.

The most common examples of thrust faulting are in the Rocky Mountains, although the structure of these mountains is so complex that it is difficult to state in simple terms how they were formed. The Colorado Rockies, for example, are primarily the result of thrust faulting and folding, produced by compressional forces in the earth's crust. Many of these faults are quite shallow, the upper portion of the fault (see Fig. 6-4b) being almost a sheet that has been moved horizontally for some distance over the lower part of the fault. These thrust sheets are so thin that they contain no metamorphic rocks. The faults in Nevada have thrown up metamorphic rocks, which are evidence that the faulting action involved rocks much farther down in the earth's crust than did the faults in the Rocky Mountains. In addition to the thrust faulting, much of the Rocky Mountain area is also folded.

6-9 Theories of Mountain Formation

Even though we might know how a particular mountain was formed, the big question still remains—what is the source of energy that produces folds and faults? In other words, what are the primal forces that are large enough to make mountains? Scientists are confronted here with a difficult

problem. Although not nearly enough information is available to solve the riddle, several possible explanations have been proposed. One theory is based on the belief that the earth is slowly cooling down and that consequently its crust is being compressed, causing folding and thrust faulting. This sounded reasonable until recent studies have revealed that the earth's temperature has been constant—kept warm probably by radioactivity—for up to two billion years, covering the time during which much diastrophism has occurred. Thus, this theory is no longer viewed as an adequate explanation for the changes that have occurred in the earth's crust.

Another theory of mountain formation, called the *convection cell theory*, is based on the heat inside the earth. In any fluid system, unless there is perfect insulation, convection currents, or movements of the fluid, transmit heat from one part of the system to another. In Fig. 6-10 a tank of water is being heated at two points. The liquid at these points is warmer and therefore has a lower density than the cooler liquid elsewhere in the tank. This difference in density causes an upward movement of the warm liquid and a downward movement of the cooler liquid in what are called convection cells, or regions of circulation. On the surface at the middle of the tank, compressional forces will concentrate because the liquid from the two convection cells meets there. With a liquid such as water these compressional forces will not be large because the viscosity, or resistance to flow, is not great. But if we liken the tank of liquid to a portion of the earth in which the convection cells are thousands of miles in size, then we have another story. On the surface of a very rigid fluid, the compressional forces would be quite large, and it has been estimated that these forces may be sufficient to produce the diastrophism that has made our mountains. Proponents of this theory believe that those parts of the earth involved in such a convection cell move only a few centimeters each century. Over an extended period of time, however, the earth's crust would become crumpled (faulted) or wrinkled (folded) depending upon the nature of the materials, the speed with which the compressional forces act, and the temperature.

Fig. 6-10. Cross-section of a tank of liquid showing the formation of convection cells.

Figure 6-11 shows a working model of the convection cell theory. Two cylinders covered with a smooth layer of plastic

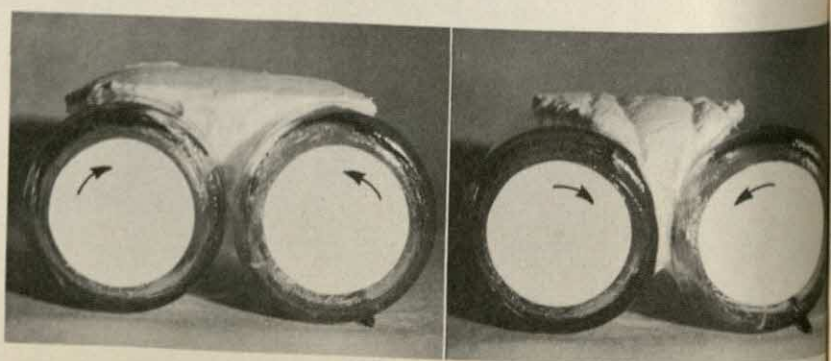


Fig. 6-11. Model of the plastic flow of the earth's crust in the convection cell theory of mountain formation. Note the upheavals produced in the originally smooth plastic surface by the rotating cylinders.

material, which is comparable to a layer of sedimentary rock in a geosyncline, are rotated toward each other, thus setting up a situation like that produced by the convection cells in the preceding figure. As the cylinders are rotated, the plastic material is compressed by lateral forces and piled up into ridges or folds whose surface is not unlike the ridges or folds that occur in nature. The convection cell theory is based on the existence of warmer and cooler areas inside the earth's mantle, a characteristic of the mantle that has never been detected by seismographic observations. In other words, the convection cell theory does not fully agree with the information we have concerning the composition and behavior of the interior of the earth.

Since almost all mountain ranges have been forced upward from geosynclines, we might re-examine a geosyncline in our search for a possible answer to the mountain-building puzzle. The convection cell theory includes a geosyncline in the center of Fig. 6-10, but there is some question regarding the source of energy that powers the whole process. In a geosyncline, there may be enough power derived directly from the force of gravity to elevate mountains to great heights. In the formation of a geosyncline, the weight of new sediments forces the old sediments downward, and the general structure after a long time shows a series of downwarped sedimentary layers that in some instances extend to a depth of 40,000 feet or more. These layers do not take on a flat shape, but are gradually downwarped, fitting into the bottom of the trough as more layers pile on top. The layers are thickest near the center and thinnest near the edge of the trough (Fig. 6-12). In effect, as a layer of rock is pushed downward a force of tension is built up in it at the same time that a force of compression builds up in the upper layers. These are action and reaction forces, initially set up by the force of gravity pulling downward upon the accumulated sediments in the geosyncline.

Reference to Fig. 6-12 will show how these forces operate. If a sheaf of papers is held firmly at the edges and then slowly bent downward in the middle, the lower papers are stretched while the upper ones are compressed or forced into wrinkles or folds. This is a model of the tension-compression theory of mountain formation, but several questions must be answered before such a theory is fully satisfactory. Are the rock layers on the bottom of the geosyncline strong enough to withstand the tension forces that would be needed to cause the upper layers to undergo folding? If not and a lower layer snaps apart under the tension force, a normal fault occurs, causing an earthquake deep down in the earth's crust. Are these tension-compression forces great enough to force parts of the upper earth's crust into folds and faults? At the present we do not know, but since almost all mountain ranges rise up from geosynclines, it is not unlikely that mountains

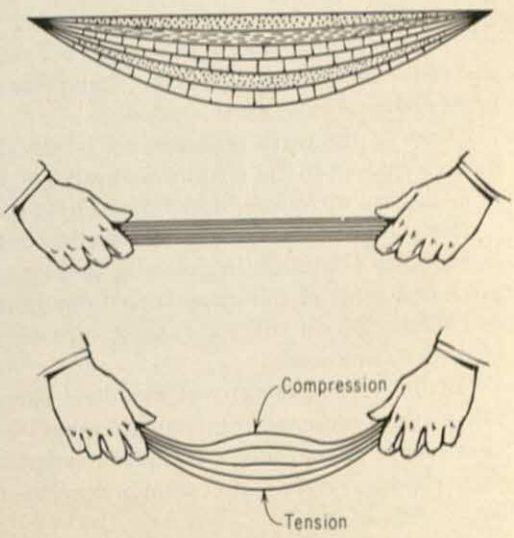


Fig. 6-12. The compression-tension theory of geosynclinal mountain formation.

are produced by forces that are inherent within the geosyncline itself.

6-10 Summary

In this chapter we have seen how the fundamental laws of force and motion operate on a grand scale in the natural world.

A warp is either a gentle downward or upward displacement of the earth's crust that extends over several miles. Downward warps are called basins, upward warps are called domes.

Folds are wrinkles produced in the earth's crust by compressional forces. The upper part of a fold is an anticline and the lower part is a syncline. In newer mountains, the anticlines are often visible, but in older mountains regions they are usually worn down by erosion until sometimes even the syncline is exposed.

Related to the syncline in shape, though on a much larger scale, is the geosyncline, which is an accumulation of sediments that forces the earth's crust downward into a huge trough. Most mountain ranges evidently are formed from geosynclines.

Faults and joints are abrupt ruptures of the earth's crust. They are produced by compressional or tensional forces that operate vertically, horizontally, or at other angles. Depending upon the forces involved, faults may be normal, thrust, or strike-slip; in each case they are evidences of great forces in operation. A joint is caused by tension forces that literally pull the rocks apart.

Earthquakes are the result of faults that occur either at the surface or deep within the earth's crust. Studies of the shock waves produced by earthquakes have given us a picture of the interior structure of the earth. Under a relatively thin crust is the mantle, and at the center is a core. The composition of the mantle and of the core, though far from certain, has been deduced from these studies.

Forces in the earth are sufficient to modify rocks through metamorphism. In this chapter we examined briefly one group of metamorphic rocks—those formed from sedimentary rocks.

Mountains are formed by erosion, accumulation, and diastrophism. Diastrophism, such as folding and faulting, has produced most of our mountains from geosyncline regions. In Chapter 24 we will see another type of mountain—those formed by volcanoes.

Finally, several theories of mountain formation were considered, and some typical mountain ranges were described as examples of them. Since it is difficult experimentally to verify these theories, there is no common agreement on their validity.

EXERCISES

A. KEY TERMS TO REMEMBER

anthracite
anticline
basin
convection cell theory
core
crust
diastrophism
earthquake
geosyncline
fold
joint
law of original horizontality
mantle
marble
metamorphism
normal fault
quartzite
slate
seismograph
strike-slip fault
syncline
tension-compression theory
thrust fault
warp

B. QUESTIONS ABOUT FORCES AND MOTION IN NATURAL PROCESSES

1. If you discover a layer of sedimentary rock that is horizontal, can you be absolutely sure that it has never been disturbed since it was originally formed?

2. Why is it more difficult to detect and study a syncline than an anticline?

3. Do warps occur in mountain regions? If so, is it easier or more difficult to find them there than on the plains?

4. Describe the difference between warping and folding.

5. Suppose a flat region near the sea underwent severe folding. Would you expect the coast line to advance landward or seaward?

6. If a syncline becomes filled with a thick layer of sedimentary deposits, what happens to the elevation of the original syncline?

7. What is the difference between a syncline and a geosyncline?

8. Which of the two fault lines in Fig. 6-5 probably occurred first? Is enough information available in this illustration to classify the type of fault?

9. Which kind of coastline would probably provide better seaports—a coastline that is advancing landward or one advancing seaward?

10. Which kind of coastlines, advancing landward or seaward, would probably produce (a) a beach, ideal for auto racing, (b) the most scenic type of coast (c) a wide, sandy beach but edged with cliffs?

11. From an inspection of the Pacific Coast of the United States, would you say that the land along the coast is rising or falling?

12. If the isostasy principle holds, would more or less rock than predicted by ordinary surveying methods need to be removed in making a highway cut through a ridge?

13. An earthquake may topple large, well-constructed buildings and at the same time leave small, flimsy structures unaffected. Why?

14. Suppose you were sitting quietly and noted an earth tremor that was followed by another tremor about 4 seconds later. At what distance would you estimate an earthquake had occurred?

15. Which type of folded mountain would probably be more accessible for geological study: a young and uneroded one, a partially eroded one, or a fully eroded one?

16. Make a list of the theories of mountain formation and show the following: the basic concept involved, the source of energy that makes the mountains, and the agreement with observations. Which theory appears most reasonable to you? What additional information would be needed to test fully your favorite theory?

17. At the end of Chapter 2 you wrote definitions of mass, space, time, and energy. Review your definitions and re-define these terms in view of your study of the last few chapters.

Seven

The Solar System

Let us now turn our attention from the earth to a different portion of the universe—the space around the earth we can explore with the naked eye. At one time or another, most of us have gazed skyward at night and wondered about the host of stars that dot the heavens. We have noticed the moon grow brighter from night to night and then wane into darkness. Against the backdrop of the stars, we have seen other bright heavenly bodies, the planets, move somewhat irregularly across the sky. Occasionally, we have been startled by the bright streak of a meteor which flashes by as if a giant had suddenly scratched a luminous mark upon the heavens.

For centuries, the natural curiosity of man has caused him to ask such questions as: Why does the sun rise and set? Why do all the planets revolve in the same direction? What causes the change of seasons? Attempts to answer these and similar questions have led man through a long and difficult search for suitable explanations. By examining the development of these explanations, we can find out not only how our present conception of the solar system was derived, but also how the scientific method operates. Keep in mind that our present notions are not necessarily correct, and that newer and more accurate theories are likely to be proposed in the future. A study of the solar system, especially after our experience over the centuries, should demonstrate that in science we may never have the final answer to a problem and should make us think twice about our modern tendency to think smugly that we have attained perfection in our theories and laws.

7-1 Early Greek Concepts of the Solar System

It is strange that early man's curiosity was attracted to the sun, moon, and stars, and that quite early in the history of science, natural philosophers applied their intellectual efforts toward explaining the behavior of these celestial bodies. The earliest concept that merits our attention here stems from the Greek philosophers in the fourth century B.C. At this time it was quite well established, through direct visual observation, that the "fixed" stars were located on the surface of a large sphere inside of which moved the planets, the sun, and the moon. The question of the relative motions of these bodies with respect to the earth constituted one of the most challenging problems to the Greek philosophers.

From our modern point of view we may be prone to underrate the ability and intellectual integrity of these early natural scientists, but within the frame of reference in which they lived, their contribution was indeed a significant step. These early workers were not equipped with an understanding of the scientific method as we know it, but rather their approach was of philosophic, or metaphysical nature. For example, Plato noted the frequent circular motion of the sun and stars as they moved with respect to the earth. Consequently, he proposed that all bodies in the sky move in circular paths and that the planets (from the Greek for wanderer) moved in complex circular paths. When a piece of information did not fit the ideal picture of circular motion, the information rather than the theory, was questioned.

One solution to Plato's problem devised by his student Eudoxos, required 26 complex circular motions to describe the movements of the planets. Aristotle added another 29 circular motions to explain some of the details Eudoxos missed, yet even in his system there were glaring omissions. It was known, for example, that some of the planets periodically seemed to recede from and advance toward the earth. Aristotle discounted this observation as insignificant, although later, of course, it was found to invalidate the whole scheme.

Another group of Greek scholars suggested that the sun is the center of the solar system. This heliocentric or sun-centered theory accounted for the receding and advancing positions of the planets, but most of the contemporaries of these imaginative observers were unwilling to accept it because the earth, the home of man, was considered the natural center of the universe—it was inconceivable to most of the men of that time that anything but the world could be the center of everything. In addition, the followers of Aristotle cleverly argued that if the earth moved, as in Fig. 7-1, then the position of the stars would be slightly altered from season to season. That is, when a near star, at point *A* in the figure, is observed in the summer it will appear against the backdrop of the stars at *C*. The same star observed in winter, however, will appear against a different group of stars. This shift is called the *annual parallax*, but the Greeks were unable to ob-

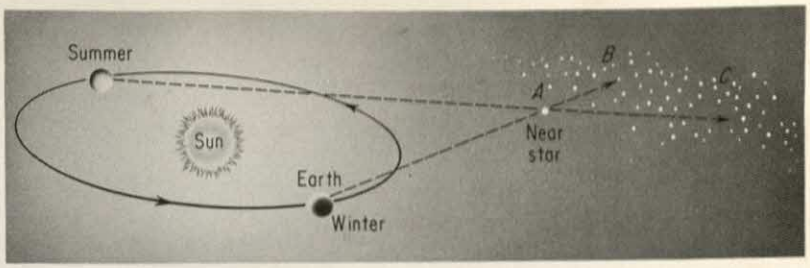


Fig. 7-1. The annual parallax of the stars in a heliocentric system.

serve it and they refused to accept the heliocentric theory. It was not until 1838 that, with the aid of astronomical telescopes, the annual parallax was observed and measured. Scientists during the Renaissance were able, through more sophisticated reasoning and with more accurate observations than the Greeks, to prove the validity of the heliocentric theory.

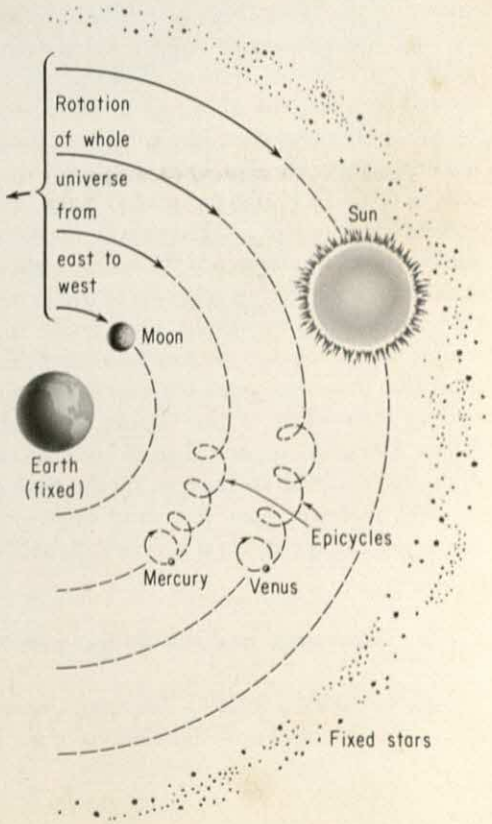
Let us further examine the geocentric or earth-centered theory before turning to the heliocentric theory because here was an idea that held the mind and interests of man for over a thousand years. Byzantine and Arabic astronomers knew of it, and it became an important element in European thought, particularly during the latter part of the Middle Ages. Furthermore, the gradual evolution of ideas about the solar system is indeed an interesting case study of how a scientific theory is developed.

7-2 The Ptolemaic Universe

Claudius Ptolemy, a geographer and astronomer of Alexandria in the second century A.D., is responsible for bringing the geocentric theory of the solar system into popular acceptance. In his theory, he attributed the rising and setting of the sun and the moon to the circular orbits of these bodies around the earth (Fig. 7-2). The sun, he explained, travels around the earth once in 24 hours, the moon once in almost 25 hours. Since the movements of the planets are more complex, they required a more elaborate model to describe their paths around the earth. If we were to observe the planet Mars over a period of time, as Ptolemy did, we would see that from night to night it moves in an easterly direction in relation to the stars in the background, but after a while, it slows down in its easterly motion, turns to the west briefly, and then returns to its original easterly movement (Fig. 7-3). Ptolemy tried to account for this retrograde movement of the planets to the west by saying that in addition to their regular circular orbit about the earth, the planets simultaneously revolve in a much smaller circular orbit, or epicycle, of their own (Fig. 7-2). By selecting the proper rates of revolution in these two orbits, Ptolemy convincingly explained the retrograde motion of the planets. It is not difficult to see that if a planet were to move in this fashion it would appear to move forward, then periodically fall behind for a bit, and then advance forward again.

The Ptolemaic theory also accounted for the fact that the

Fig. 7-2. The geocentric theory of Ptolemy.



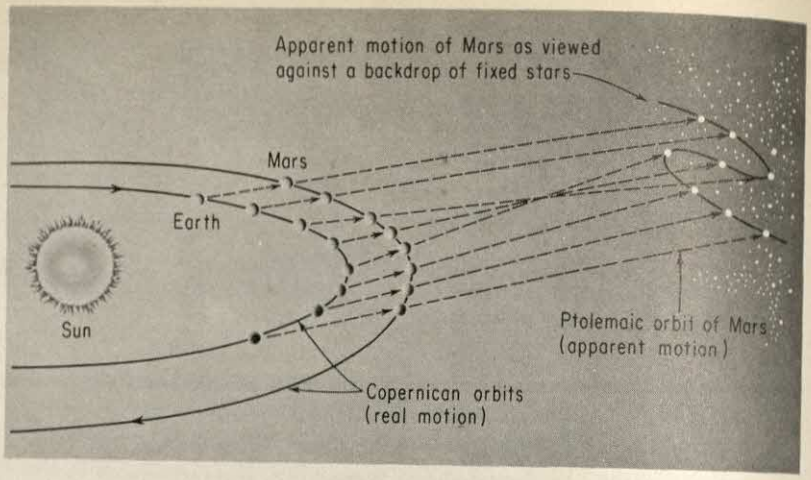


Fig. 7-3. The apparent retrograde motion of Mars as explained first by Ptolemy and later by Copernicus.

sun as well as the planets vary in their distances from the earth. For instance, the sun is noticeably—three million miles—nearer the earth in winter than it is in summer. Why then, you may wonder, is it not hotter in winter than it is in summer? The reason is that the earth's axis is at an angle which causes the Northern Hemisphere to be tilted toward the sun in summer and away from the sun in winter. The sun's rays, consequently, strike the Northern Hemisphere at a greater slant in winter than in summer, and their angle to the earth affects the seasons more than the varying distance between us and the sun. Moreover, since the average distance from the earth to the sun is about 93 million miles, the 3-million-mile variance is not a large one relative to the whole distance.

In summary, Ptolemy's theory held that the heavens are a huge sphere, and at the center of this sphere is an immovable earth, around which revolve the sun and the planets in a complex combination of at least 70 circular motions.

The Ptolemaic theory explained the facts as they were known at that time. It predicted the movements of the planets and other heavenly bodies accurately enough for navigation purposes. It showed that the stars should exhibit no annual parallax, and above all it was accepted and approved by the contemporary religious leaders. Furthermore, we must not lose sight of the fact that a geocentric model of the solar system appeals to "common sense." Even today we commonly speak of the sunrise and the sunset, implying that it is the sun that moves with respect to the earth. Although common sense is at times a dangerous criterion to use in science, it has served effectively countless times as a guide in the search for a simple and adequate answer to a scientific problem. For almost two thousand years the geocentric concept was the prevailing view—first of the Greeks, then of Ptolemy and his fellow Alexandrians, and then of both the eastern and western divisions of Christian Europe. By the time of the Renaissance in 1500, however, new data and more penetrating analyses subjected the theory to a severe rational test.

7-3 Copernicus and the Heliocentric Theory

On his deathbed in 1543, Nikolaus Copernicus saw the first copy of his famous book, *Revolutions of the Heavenly Spheres,*

in which he proposed that the sun, not the earth, is the center of the solar system. In a way he was fortunate that the book had not been published sooner, because the geocentric theory had become so thoroughly established that it was dangerous to publicly oppose it. For many years after the Copernican system was first proposed, advocates of it were persecuted. Bruno, an Italian philosopher and a contemporary advocate of Copernicus' theory, was burned at the stake in 1600 partly for his outspoken support of the revolutionary theory.

Copernicus was one of the men who created the Renaissance, along with, to mention just a few, Columbus, Vasco da Gama, and Magellan in geographical exploration, Michelangelo and da Vinci in art, Luther and Zwingli in religion. He lived in a time of artistic expression and of commercial and military expansion, and it was in this lively period that modern science was born. Although trained in medicine and economics, Copernicus became intensely interested in the structure of the solar system. He was a good mathematician, and before he died he had developed his theory to an extent that eventually caused it to be accepted as superior to the Ptolemaic theory in its explanation of the structure and behavior of the sun's family. He failed to explain many things, but the general framework of his theory was sound, especially after Kepler replaced circular orbits with elliptical orbits in experiments that we will consider in detail later. Copernicus, in short, forced the world to take probably its greatest single leap forward in its knowledge of the universe.

The Copernican theory set the sun at the center of the solar system (Fig. 7-4) and said that the earth and the other planets (six were known at the time) travel around it in circular paths, a proposition that greatly simplified the Ptolemaic view of how these various bodies move. The sun remains fixed in position with respect to the circular orbits of the planets, although Copernicus had to assume that the sun is slightly removed from the center in order to account for the receding and advancing position of the earth from summer to winter. In contrast to the Ptolemaic system, the Copernican theory held that the earth rotates on its axis and that this spinning action produces the effect of a rising and setting sun. The new theory, however, still considered the stars to be fixed in a huge sphere in much the same manner as that proposed by Ptolemy.

In the Copernican theory, the retrograde motion of the planet Mars is explained by the fact that both Mars and the earth move in the same direction in their orbits around the sun (Fig. 7-3). Since the earth is closer to the sun than Mars, it will be accelerated around the sun at a more rapid rate than Mars—the earth revolves once in 12 months, while Mars requires 22.6 months to make the circuit. Consequently, the earth periodically will catch up with Mars and pass it. To an observer on the earth, as we have seen, Mars appears to stop in its usual easterly motion, and, while the earth passes

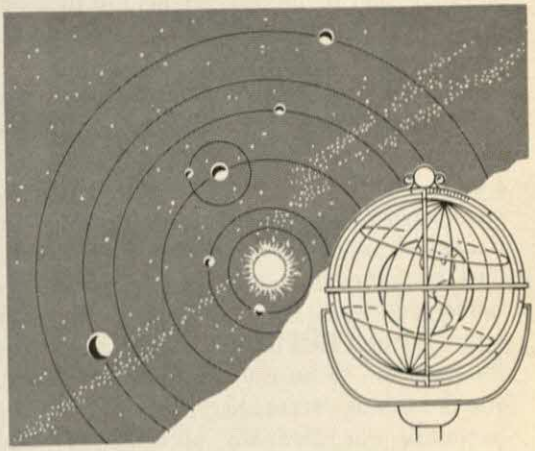


Fig. 7-4. The Copernican heliocentric system, with only the planets known to Copernicus, and the famous "Jagiellonian golden glove," constructed about 1510 and used in his studies of planetary motion. (Redrawn from Stephen P. Mizwa, *Nicholas Copernicus,* The Kosciusko Foundation)

it, appears to move toward the west. We note the same effect of relative motions when we pass another car on a highway. Before and after we pass the car, it appears to be traveling in the same direction as we are. But during the few moments we are side by side with the other car, it appears to be moving in the opposite direction. This explanation of the retrograde movement of the planets was, because of its simplicity, much more appealing than Ptolemy's complicated system of epicycles.

The opponents of the Copernican theory, which at first included nearly everyone, tried to discredit Copernicus by asking him embarrassing questions. One of the first things they wanted to know was why the earth, if it revolves around the sun, does not fly away from the sun, much as mud flies off a rapidly turning wheel. To this, Copernicus replied that the bodies in his system were not likely to fly apart because they did not move as fast as the ones in the geocentric system, in which the sun circles the earth every day. Today, of course, we know that the planets are held in their orbits by the force of gravity, but, at that time, the work of Galileo and Newton was yet to come.

The disbelievers also naturally asked why there was no parallax among the stars (Fig. 7-1). This was a tough one for Copernicus, but he suggested that the stars are much farther away from the solar system than originally believed, so far away that the best instrument of the day could not detect the parallax. This argument convinced few of Copernicus' contemporaries because they could not conceive of the universe as being so large. Furthermore, with Heaven in the starry sky and Hell beneath the earth, the faithful of the time did not care to be isolated even more from a Heaven that was becoming increasingly distant. The Copernican theory, however, was gradually accepted as valid, even by the Church, primarily because of its simplicity. In contrast to over 70 combinations of circular orbits in the geocentric theory, the heliocentric theory required only 34. Although both theories accurately predicted the behavior of the planets, science, which tends to seek the simplest answer to a problem, chose the simpler view as soon as human prejudices were overcome.

7-4 Theory in Physical Science

Now is a good time to re-examine the role of theory in physical science. In Chapter 1 we surveyed the scientific method and found that one important feature of it is the formulation of a theory. The primary purpose of a good theory is to enable us to see a natural phenomenon as part of a simple, unified whole. Even though our observations may reveal a certain number of facts, and these facts may be summarized into a law, there remains this unique feature of the

scientific method to help us gain an insight into why and how the phenomenon exists and operates.

Let us list briefly the requirements of a good scientific theory and then see how the Ptolmaic and Copernican theories compare in the light of these criteria. An acceptable theory should:

1. Correlate many facts in a single, unified concept.
2. Suggest new ideas and stimulate research.
3. Be useful in solving long-range problems and in making predictions that can be checked by experience.
4. Be based on as few and as reasonable assumptions as possible.
5. Be able to accommodate new ideas and concepts when necessary.

Needless to say, it is difficult to formulate a theory that meets all these requirements, and many of our scientific theories fall short of these standards. Indeed, we are fortunate when we produce a theory that fulfills all these requirements. If a theory does not explain all the facts, it must be modified or we must seek a more adequate theory, and if a theory fails to accommodate or stimulate new ideas and research, it falls into disrepute. The Ptolemaic theory, as we have seen, served for many centuries, but as more information and observations became available, it was no longer adequate, even with modifications. Hence, it had to be discarded because it was not productive as a workable theory.

Here are some of the facts that must be explained by any astronomical theory. The earth is an oblate sphere, 26.7 miles greater in diameter at the equator than at the poles, and its gravity vector is greater in magnitude at the poles than at the equator. Ocean and air currents travel in clockwise circuits in the Northern Hemisphere and in counterclockwise circuits in the Southern Hemisphere. Another fact to be accounted for is the behavior of the so-called Foucault pendulum. In 1851 Jean Foucault suspended a metal ball on a wire 200 feet long from the ceiling of the Pantheon in Paris. He set the ball into a to-and-fro swinging motion and observed that its path gradually changed with respect to the floor. With the passage of time the path of the pendulum deviated more and more from the path of its original swing, which Foucault had carefully marked. Since there was no force that could have caused the pendulum to change its path, he concluded that the change was due to the rotation of the earth beneath the pendulum. At the North Pole, the pendulum would turn through a complete circle in 24 hours—actually the pendulum would continue to swing in its original path while the earth beneath it rotated. Foucault thus provided the first direct evidence that the earth spins on its axis.

Still another fact that a theory must explain is the peculiar way the light from the stars strikes us. This peculiarity, called *aberration*, is caused by the movement of the observer, in

much the same way that a man sees raindrops seemingly slanting toward him when he runs through a rain storm. Even though the rain falls nearly vertically, the drops will seem to come at the man at an angle, and the faster he moves the greater the angle will appear to be from the vertical. If we consider the falling raindrops to be the light coming to us from a star, we have an accurate illustration of the aberration of starlight. James Bradley in 1728 observed that as the earth moves across the path of light coming from a star, the star appears to be slightly displaced from the position it has when the earth moves either directly toward or away from the star. This slight displacement is in the direction of the motion of the earth and confirms the theory that we move in an orbit around the sun. By measuring the amount of aberration, astronomers can tell us how fast we are moving. Their calculations show that we are traveling through space at the rate of 18.5 mi/sec, a speed that turns out to be exactly that required to get us around the sun in one year.

All these facts—the oblate shape of the earth, the curvature of air and ocean currents, the variation in the gravity vector, the Foucault pendulum, the annual parallax, and the aberration of starlight—are most readily explained by the concept of a rotating earth revolving around the sun. Now we can go back to the criteria of a good theory at the beginning of this section and assess the Ptolemaic and Copernican theories. The Ptolemaic theory, as we see, falls down on points 3, 4, and 5 outlined above. It did not accommodate much of the new information found in the years following the invention of the telescope, it failed in making satisfactory predictions of the movements of the planets, and it was not nearly as simple as the other theory. The Copernican theory, on the other hand, has fulfilled all five requirements. It has been notably successful in accommodating new information, particularly the discovery of several new planets that were unknown at the time of Copernicus. As we shall see in the next section, it was through the application of this theory and the laws of gravity and motion that the existence of these unknown planets was predicted. In view of our best information today, the idea that we are in a solar system with a sun at its center and with planets revolving around the sun is more adequate than any other theory that has been proposed. However, we must not think that this theory will never be changed or supplanted—it is just the best one we have with our present state of knowledge.

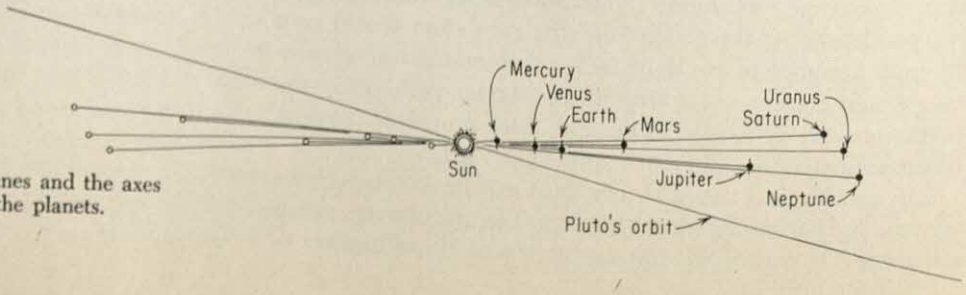


Fig. 7-5. The orbital planes and the axes of rotation of the planets.

7-5 The Planets and Their Orbits

There are nine planets in the solar system, and they all move around the sun in the same direction and in nearly the same plane as the earth revolves around the sun. To a space traveler approaching the sun, the solar system with its nine planets would appear very flat and pancake shaped (Fig. 7-5). Pluto, which is really not a full-blown planet but in a class by itself, is the most erratic of the planets, and it revolves in an orbit that it tilted 17 degrees from the plane of the earth's orbit. Furthermore, each of the planets spins on its own axis at rates varying from once in 10 hours for Jupiter to once in 88 days for Mercury. With the exception of Uranus, the axes of rotation of the planets are tipped slightly, several of them being tipped at almost the same angle as the earth —23.5 degrees. Uranus is tipped 98 degrees, with the result that there is a 42-year "summer" and a 42-year "winter" at each of its poles during the 84-year period of its orbital cycle around the sun. Uranus' five moons, incidentally, are in orbits that are also tipped about 98 degrees.

The inner planets revolve around the sun at a faster rate than do the outer planets: Mercury travels almost 30 mi/sec while Pluto creeps along at less than 3 mi/sec. Remember that the planets are constantly being accelerated toward the sun by the force of gravity, which is proportional to the product of the two masses involved and inversely proportional to the square of the distance between them. Although Pluto is several times heavier than Mercury, it is almost 100 times farther from the sun, and consequently there is a much smaller gravitational force acting on it than on Mercury. Since Mercury is accelerated toward the sun with a greater force, it must revolve at a higher rate so its centrifugal force will balance the gravitational attraction of the sun. While Pluto is revolving around the sun once, Mercury makes over a thousand revolutions.

The factors of size, shape, density, movement, and distance from the sun seem to classify the planets in two major groups —the inner four and, excepting Pluto, the outer four. The inner planets rotate on their axes much slower than do the outer planets. Mercury rotates only once during each revolution, consequently the same side is presented to the sun at all times. Since Venus is cloaked by a thick cloud we cannot determine its rate of rotation, but Mars rotates at almost the same rate as the earth. In contrast, the outer planets rotate at much greater speeds in spite of their much greater size. Three of the outer planets rotate once in approximately ten hours (see Table 7-1). For a large planet such as Jupiter, this means that the surface at its equator spins at the terrifically high rate of 28,000 mi/hr, which is 28 times the rate at our own equator. As a result, the equatorial bulge of Jupiter is much greater than that of the earth and is sufficient to cause the planet to appear noticeably flattened (Fig. 7-6).

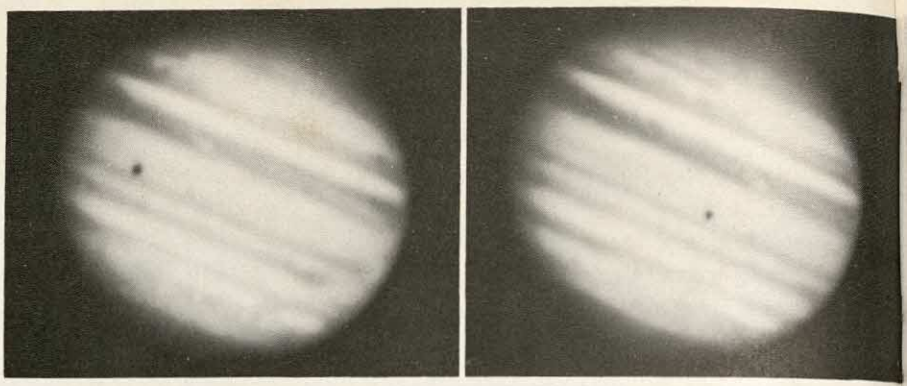


Fig. 7-6. Two photographs of Jupiter taken fifty minutes apart. Note the movement of the spot, from which the rate of rotation can be determined. (Mount Wilson and Palomar Observatories)

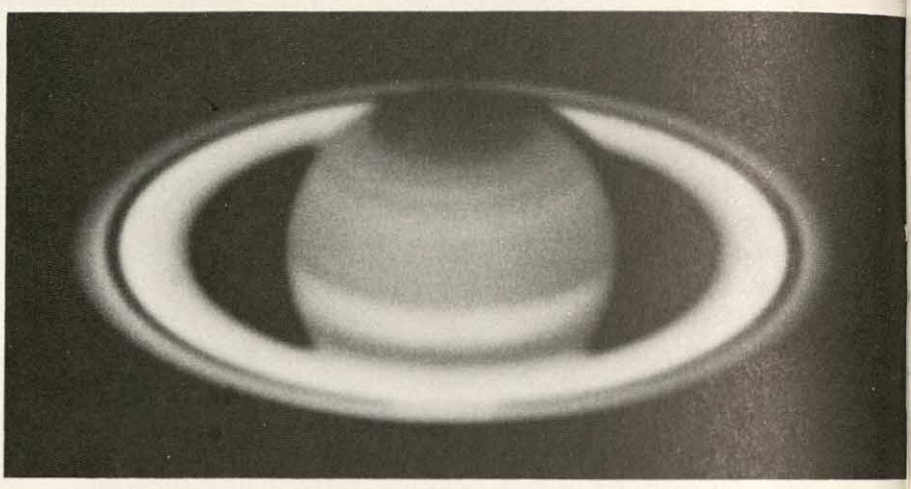


Fig. 7-7. Saturn and its ring system. (Mount Wilson and Palomar Observatories)

Saturn (Fig. 7-7), Uranus, and Neptune also have very high rotation rates and are similarly flattened.

The shapes of the planets are determined by the relationship between the centripetal force of gravity and the centrifugal force supplied by their rotation. If the planets did not rotate, they would be perfect spheres, because there would be no centrifugal force and it is the natural tendency of a large mass to assume a spherical shape unless acted upon by outside forces. Since they do rotate, however, they bulge out at their equators where the rate of rotation and therefore the centrifugal force, is greatest.

Table 7-1 gives the vital statistics of the solar system. These sizes and distances are so enormous compared to those we are familiar with that they put a great strain on our imaginations. Note, for instance, that the distance from the sun to Pluto—in other words, the radius of the solar system—is one million times greater than the width of the United States.

With the exception of Mars, the other planets are greatly different from the earth in almost every respect. Mars is intriguing because of the possibility that it supports life. It has a rather frigid, desert climate—although some observers have identified green areas separated by canals—and polar icecaps that advance and recede with the seasons. The clouds in its atmosphere move much as ours do, but its atmosphere creates a curtain through which we can see the surface of the planet only faintly, and for this reason much of the information we have about Mars is debatable and inconclusive.

Venus, the nearest planet to the earth is, next to the sun and the moon, the most brilliant object that is regularly seen

in the sky. Because the orbits of Venus and Mercury are inside the orbit of the earth, these planets appear only as evening or morning stars and then never very far above the horizon. Only when they are on the far side of the sun from the earth do we see their fully illuminated faces; the rest of the time only part of their illuminated surface is visible to us. Recent studies of the light and heat emitted by Venus indicate that the temperature of this planet may be above the boiling point of water, which is not surprising since it is so close to the sun.

The outer planets—Jupiter, Saturn, Uranus, and Neptune —are characterized by low density and great size. Jupiter's face appears to be composed of belts 10,000 to 15,000 miles wide which rotate at different velocities and indicate that the surface, at least, of the planet is largely gaseous. Saturn is unique in possessing three flat concentric rings around its equator, which vary from 1000 to 16,000 miles in width and are distinctly separate from each other. These rings rotate along with the body of the planet, although the outer parts of each rotate slightly slower than the inner parts. Bright stars shine through the rings, which we think are composed of small solid particles. Uranus has a cloudlike surface that is evidently composed of gases, and it probably has a small, solid core.

The discovery of Neptune is a dramatic example of scientific detective work, although the circumstances surrounding its discovery are clouded by the foibles of erring man. In 1841, while still an undergraduate at St. Johns College in England, John C. Adams became interested in the fact that Uranus progressively wandered slightly from its regular orbit. It had been assumed that these perturbations were caused by another as yet unknown planet. However, the problem of reasoning out the position of the unknown planet required the most brilliant of mathematical ability. Upon graduation in 1843, Adams seriously undertook the solution of the problem and in about a year came up with a predic-

TABLE 7-1 THE SOLAR SYSTEM

	Distance from sun (mi)	*Diameter (mi)*	*Density g/cm³*	*Period of revolution*	*Period of rotation*	*Weight of 1-lb mass at surface*
Sun	——	864,000	1.4	——	25 days	27.9
Mercury	36,000,000	3,000	3.8	88 days	88 days	0.3
Venus	67,000,000	7,575	5.1	225 days	?	0.9
Earth	93,000,000	7,918	5.5	365 days	24 hr	1.0
Mars	142,000,000	4,216	4.0	687 days	24.5 hr	0.4
Jupiter	483,000,000	86,000	1.3	12 years	10 hr	2.6
Saturn	886,000,000	72,000	0.7	29.5 years	10.2 hr	1.2
Uranus	1,780,000,000	30,000	1.4	84 years	10.8 hr	1.0
Neptune	2,790,000,000	32,000	1.3	165 years	15.8 hr	1.0
Pluto	3,670,000,000	smaller than earth	?	248 years	6.39 days	?

tion of the position of the unknown planet. However, because Adams had tackled this difficult problem without assistance or guidance of any kind, there was some question about its validity. The Royal Astronomer was so insufficiently impressed that he failed to turn his telescope to that point in the night sky indicated by Adams' concise calculations—a simple observation that would have resulted in success.

Meanwhile, Leverrier, an established French astronomer, had also begun to study the perturbations of Uranus, and, independently of Adams, he arrived at a solution that was almost in exact agreement with that of Adams. Again, the Royal Astronomer had the opportunity of making the actual discovery of the new planet, but because of the rather careless observations of his assistant, who actually saw the new planet but failed to recognize it, it still remained hidden in the recesses of space. Leverrier wrote to Galle, a German astronomer, pointing out the location of the unknown planet, and Galle, with the help of a sky chart so new that the English had not yet obtained a copy, found the new planet in almost exactly the position predicted by both Leverrier and Adams. Although the actual discovery of the new planet, Neptune, was based upon Leverrier's calculations, Adams' work was completed earlier and would have resulted in the discovery had his results been properly handled. Consequently, Adams and Leverrier share equally in the credit for calculating the position of Neptune on the basis of the perturbations of Uranus.

Pluto was discovered in very much the same way as Neptune. In 1916, Percival Lowell, who established the observatory at Flagstaff, Arizona, calculated Pluto's position by noting the periodic deviation of Neptune from its regular orbit, and in 1930, Clyde Tombaugh discovered the new body by comparing a series of photographs (Fig. 7-8) taken of that point in the sky where Lowell said it should appear. However, the status of Pluto as a regular planet has always been questioned. It has an unusually eccentric orbit, the plane of which is tilted away from the other planets, and in view of its remote position at the fringe of the solar system, its small size is puzzling. One possibility is that Pluto may have once been one of Neptune's moons, since it acts more like a stray satellite than a planet.

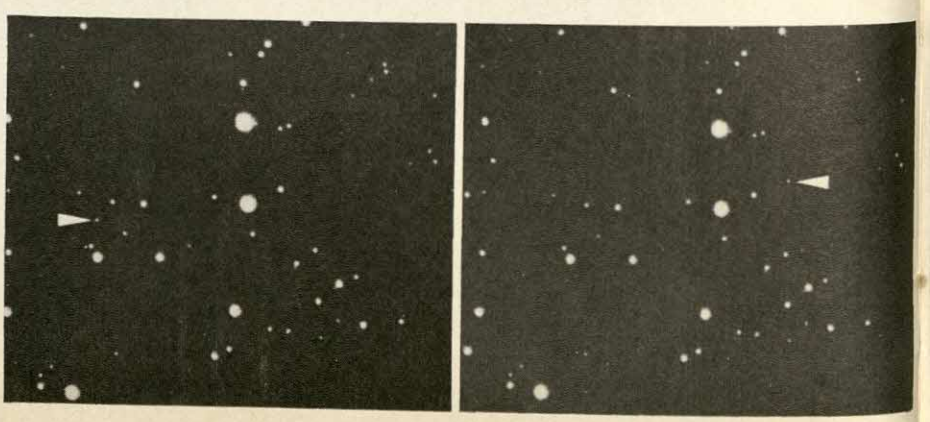


Fig. 7-8. Two photographs from which the discovery of Pluto was made in 1930 by Clyde W. Tombaugh. The arrows show the movement of the planet during an interval of six days. (Lowell Observatory photo)

7-6 Kepler's Laws

During the half-century following Copernicus, Tycho Brahe, a Dane who disappointed his noble family by studying what they considered to be the ungentlemanly science of astronomy, collected an astonishing amount of accurate data concerning the solar system. Although he rejected the Copernican system because he could not observe the annual parallax with his naked eye, his careful observations of the movements of the planets provided the information that enabled his student, Johannes Kepler, to formulate three very important laws of motion that apply to the solar system.

Kepler was a patient, tireless worker who spent almost a lifetime analyzing Tycho's data. He found that these observations of planetary motion fitted into neither the old Ptolemaic theory nor the simpler Copernican theory. Confident that the data were accurate, he conceived and applied them to a new heliocentric theory that nicely explained Tycho's data. The basis of Kepler's new concepts are embodied in three laws of planetary motion which have become the backbone of modern astronomical science. The work of these two —Tycho and Kepler—represents the ideal combination of precise observation and thorough interpretation, two complementary processes that have often brought outstanding achievements in science. Let us now examine these three laws of planetary motion.

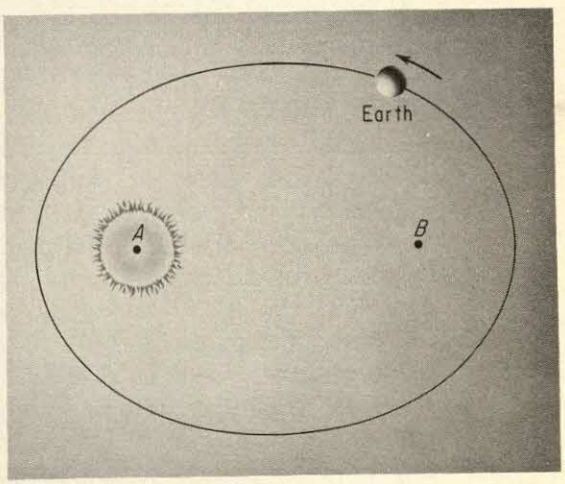


Fig. 7-9. An ellipse with the sun at one of the foci.

1. *The orbit of each planet is an ellipse with the sun at one of the foci* (Fig. 7-9).
   An ellipse is the locus of points the sum of whose distances from two foci is a constant. The further the two foci are from each other the greater is the eccentricity of the ellipse. When the two foci are at the same point, the ellipse is a perfect circle. An elliptical path for the planets eliminates the need of superimposing other motions on a circular orbit, and it neatly accounts for the fact that the earth is closer to the sun in winter (91,300,000 mi) than in summer (94,500,000). Kepler repudiated the classic concept that all celestial motions are circular, yet he probably wondered why the planets have elliptical orbits—what is there about an ellipse that is more fundamental than a circle in the motion of the planets? An answer to this question came eighty years later when Newton concluded from his laws of motion that the elliptical path is a natural consequence of the gravitational forces that hold the planets in their orbits —it is more or less an accident if a planet follows a perfectly circular orbit and even today, the modern space technologist celebrates the occasion on which a man-made satellite is placed in exactly a circular orbit around the earth. We will now see how these gravitational forces are involved in Kepler's second law.
2. *A line joining the planet and the sun sweeps over equal*

areas in equal time intervals, regardless of the length of the line (Fig. 7-10). Kepler noted in Tycho's data that the planets appear to move more rapidly during some phases of their orbital journey than at others. But how could these data be fitted into a logical model? The planets could hardly be expected to speed up and slow down of their own volition. After voluminous calculations, Kepler was overjoyed to find that the answer was the simple and direct relationship stated in the second law. He discovered that the earth travels most rapidly in midwinter and slowest in midsummer, and that this periodic speeding up and slowing down accounts for its elliptical orbit. But what accounts for these changes in speed?

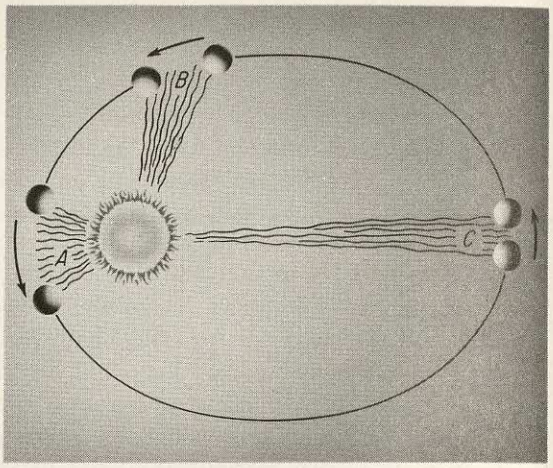
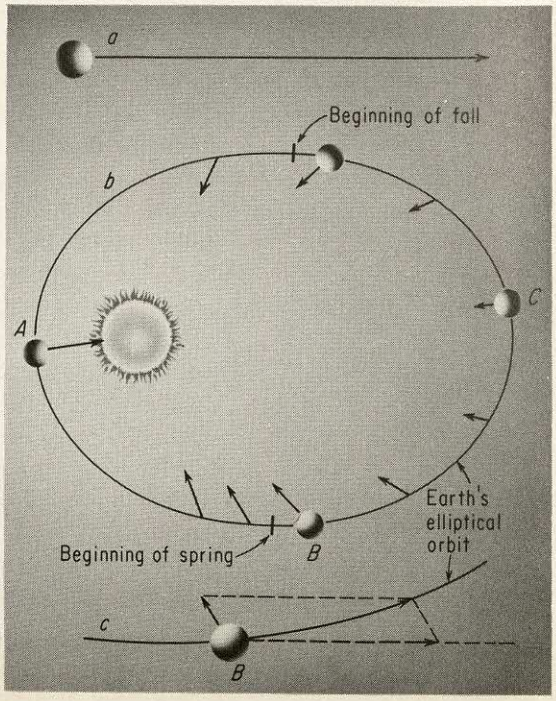


Fig. 7-10. Kepler's second law. The three areas swept out in equal time intervals are equal to each other.

You will recall that Newton's first law says that an object in motion will tend to remain in uniform motion unless acted upon by an external force. Thus, in Fig. 7-11a, if the earth were moving and no force were acting upon it, it would move uniformly in a straight line. In Fig. 7-11b the centripetal force of gravity between the sun and the earth causes the earth to follow a curved path. Note that the gravity vector is almost at a right angle to the earth's orbit at *A*, but that between *A* and *C* the gravity vector is at a backward angle, which tends to slow down the orbital velocity of the earth. Figure 7-11c shows the component vector forces operating on the earth at a point between *A* and *C*. After the earth passes *C*, the gravity vector is at a forward angle, thus causing the earth's orbital velocity to be increased. Finally, at *A* the gravity vector is once again at a right angle to the orbit. The changing angle of the gravity vector thus accounts for the periodic speeding up and slowing down of the earth, causing it to take 186 days to travel in its orbit from the beginning of spring to the beginning of fall and only 179 days for the other half of the trip. Since, in an elliptical orbit, the gravity vector is greatest in magnitude when the planet comes nearest the sun, as at *A* in Fig. 7-11b, the earth experiences its greatest acceleration at this point and will swing around the sun more rapidly at *A* than it does at *C*. Newton therefore, took Kepler's second law and demonstrated that the ellipse is the most natural orbit for a planet to follow. Now for Kepler's third law.

Fig. 7-11. The changing directions of the forces that act on the earth in its elliptical orbit.



3. *The square of the time of a planet's revolution is proportional to the cube of its average distance from the sun.* Mathematically, this law can be stated in a simple equation where T is the time required for a revolution and D is the distance between the planet and the sun:

$$T^2 \propto D^3$$

or:

$$T^2 = kD^3 \qquad (7.1)$$

The value of the constant k is the same for *all* planets. This law makes it possible to calculate either the period of revolution or the distance of the planet from the sun when one or the other of these factors is known. For example, a planet that is 9 times as far from the sun as the earth must have a period of revolution 27 times as long. An accurate check on the validity of this law can be made by comparing the movements of the earth and Neptune. Fitting the data in Table 7-1 into a modified form of Equation 7.1 we have:

$$\frac{(\text{Earth's period})^2}{(\text{Neptune's period})^2} = \frac{(\text{Earth's distance})^3}{(\text{Neptune's distance})^3}$$

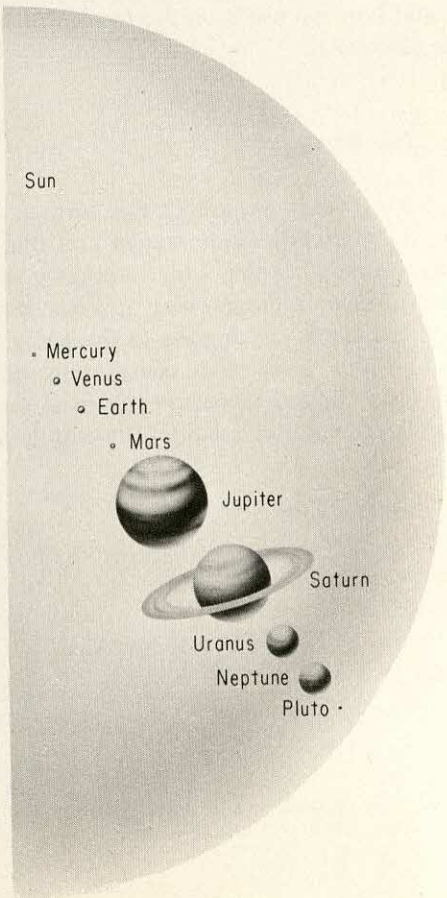
$$\frac{(1\text{ yr})^2}{(165\text{ yr})^2} = \frac{(93{,}000{,}000\text{ mi})^3}{(2{,}790{,}000{,}000\text{ mi})^3}$$

$$\frac{1}{27{,}225} = \frac{1}{27{,}000}$$

The small difference between these two numbers reflects the fact that the data in Table 7-1 have been rounded off. With more precise data these two numbers would be practically equal to each other.

Kepler's laws have been checked by our modern instruments and telescopes and found to be extremely accurate. In place of the complicated Ptolemaic system of epicycles, he gave us a remarkably simple view of the motions of the planets and a logical model of the solar system that can be described by relatively simple mathematical equations.

Fig. 7-12. The relative sizes of the sun and the planets.



7-7 The Sun

By far the largest member of the solar system (Fig. 7-12), the sun is not only the central body but the only luminous one in the system. The planets merely reflect the light of the sun, which is a star of average brightness and size. It appears much larger and brighter than other stars, of course, because it is much closer to the earth. Galileo, with the use of his early telescope, was the first to discover that the sun has spots on it. Ever since, sunspots have been of major interest to the astronomer and have provided a direct method of determining the sun's period of rotation. These spots are dark areas on the surface of the sun that may be as large as 100,000 miles in diameter (Fig. 7-13). Their size and shape vary from day to day, and for some unexplained reason, they appear most commonly near the sun's equator.

The sun is a body of luminous gases that rotates on its own axis once each 25 days. Its surface temperature is about 6000°C, and its interior temperatures have been estimated at 25 million degrees. We shall find in later chapters that the sun is a star, and a rather ordinary one at that, which gives off prodigious amounts of energy in the form of heat and

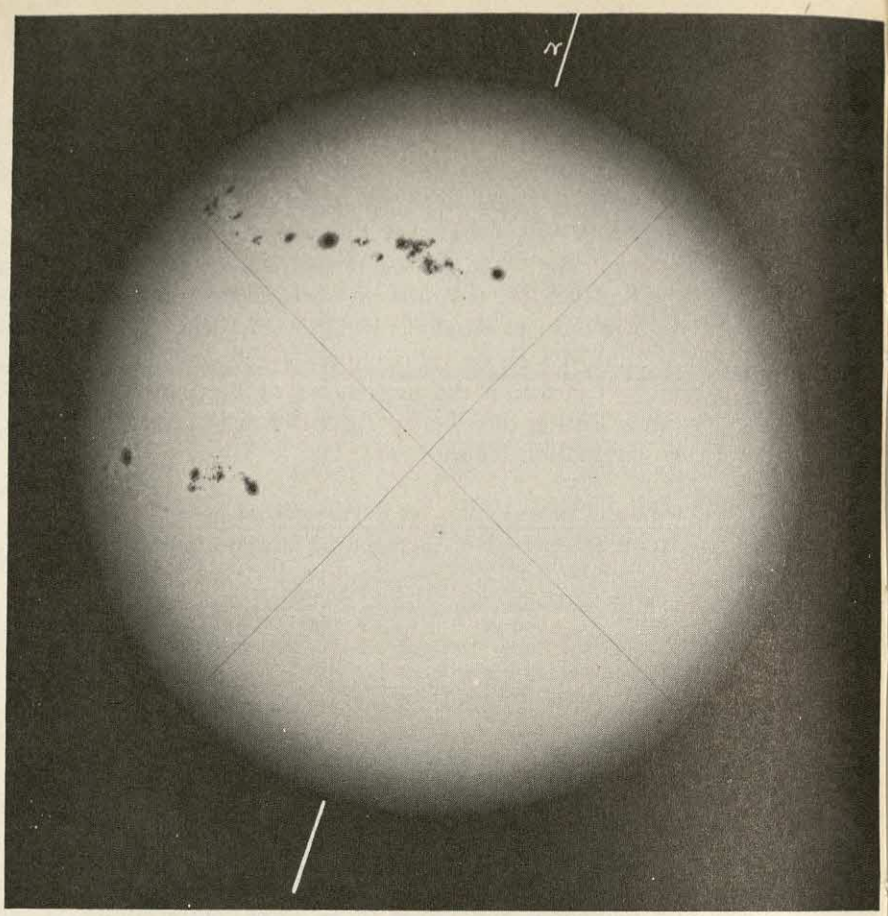


Fig. 7-13. Sunspots on the surface of the sun, February 1956. (From Royal Greenwich Observatory, copyright reserved)

light. The way this energy is produced, how it is transmitted to us, and how we use it on the earth will be brought out in a later discussion.

7-8 The Earth

We have already seen that the earth is one of the nine known planets in the solar system and that it rotates on its axis once each 24 hours. This rotation gives us the illusion that the sun rises in the morning and sets at night. The earth's axis is tilted almost 66 degrees to the plane of its orbit (Fig. 7-14), and this inclination causes our seasons. When the North Pole is tipped farthest away from the sun, about December 21 (the winter solstice), the sun is directly overhead

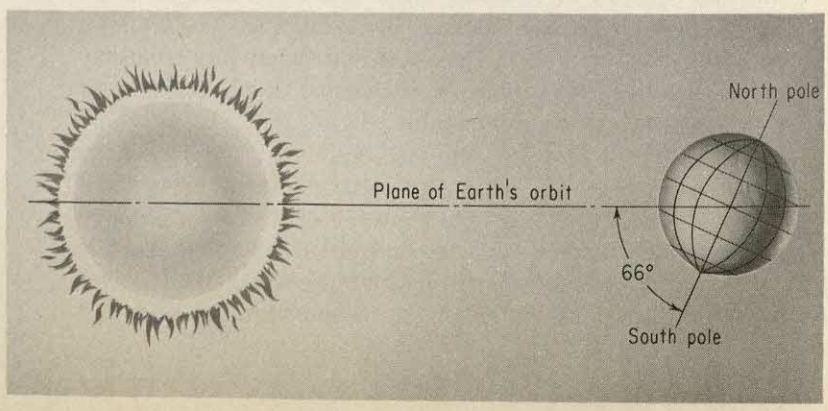


Fig. 7-14. The earth's axis is inclined to the plane of the earth's orbit.

at the latitude of the Tropic of Capricorn, nearly 2000 miles south of the equator. Six months later, when the earth is on the opposite side of its orbit, the North Pole is tipped toward the sun and it enjoys 24 hours of daylight each day. Near March 21 and September 23—the spring and autumnal equinox (equal night and day), respectively, the sun is directly overhead at the equator at noon, and the two poles are at the very edge of the sun's illumination.

As the earth revolves around the sun, the pattern of stars in the night sky gradually changes. From night to night the stars appear more and more to the west as new groups of stars come into view at the eastern horizon. This celestial parade passes by once each year, during which time we can look out into all corners of the sky, except that part above the South Pole. If we lived at the equator, however, we would be able to see all the stars in the sky during the course of a year.

7-9 Moons and Satellites

Around most of the planets in the solar system revolve small bodies called moons or satellites. The earth has a single moon whose orbit is less than half a million miles in diameter. The moon itself is 2160 miles in diameter, or about one-quarter that of the earth. Its density is 3.5 g/cm^3, or about 0.6 the earth's density. Since the moon rotates on its axis exactly once each time it revolves in its orbit, only one side of it is ever directly visible to us. We will have to rely on a space vehicle going around the backside of the moon and sending us television pictures much like the closeup views of the visible side of the moon we have already seen.

The moon has no atmosphere, no water, and apparently no life as we know it. Its surface is composed of a variety of land forms—jagged craters, mountain ranges, and smooth areas called *mares* or seas that are apparently covered with dust or fine particles. Most of the larger craters and the flat, smooth areas have even been named. The mountain ranges vary in height from an average of 5000 feet to a maximum of nearly 27,000 feet. The craters are approximately 3 to 10 miles in diameter, although some of the mountain-ringed areas, popularly called craters, are upward of 140 miles in diameter. There is considerable variation in color on the moon's surface, with yellows and grays predominating, but greens, browns, and full-reddish colors are also common.

How the craters on the moon were formed has long intrigued astronomers. One of the most plausible theories is based on the concept of mechanical energy with which we are already acquainted. A meteor, moving at a typical speed of 40 mi/sec, if suddenly stopped would release 400 times the energy of an equivalent weight of dynamite. That is, a 1-pound meteor would release 400 times as much energy when it strikes a surface as does a 1-pound charge of dynamite

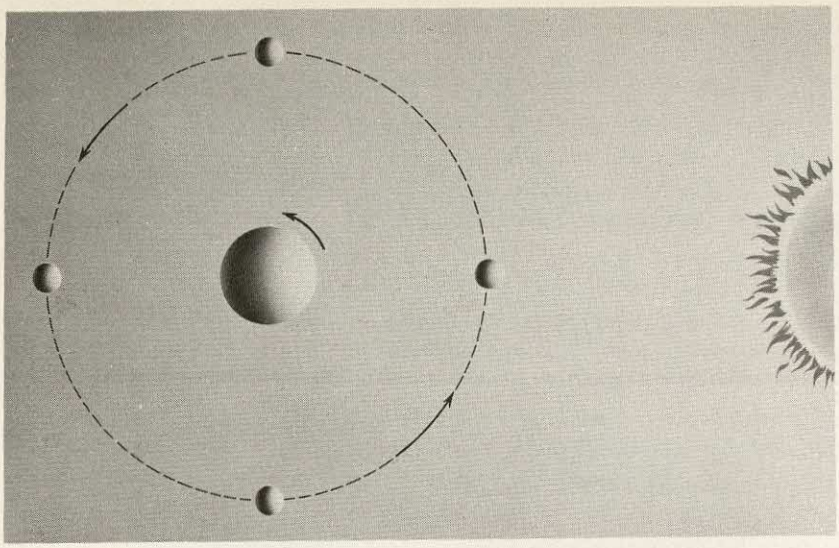


Fig. 7-15. Phases of the moon.

when set off. This abrupt conversion of kinetic energy into heat energy would be sufficient to heat a meteor and the surface of the moon to a temperature high enough literally to vaporize the meteor and part of the moon's surface. A tremendous explosion would ensue that could produce a hole possibly 2 miles deep and thus form a crater.

The moon encircles the earth once each month, traveling in the same direction as the earth rotates on its axis (Fig. 7-15). Since the moon travels approximately $\frac{1}{30}$ of the way around the earth in one 24-hour day, it will appear to rise almost an hour later each day. Let us follow the moon for a month in its journey around the earth. When the moon is located between the earth and the sun, we have what is called a new moon (Fig. 7-15). A day or so later, shortly after sundown we will see the crescent moon very low in the western sky. The diagram shows why we see only a small part of the moon illuminated; most of its surface facing us is in shadow. From night to night the crescent grows larger and moves higher in the sky. As the moon rises farther to the east each night, it appears half, then fully illuminated, and when completely full, rises in the east at the same time the sun sets in the west. After that, the moon wanes in brightness and, rising later each evening, goes to its last quarter and finally back to its original new-moon phase. When the moon is between its last quarter and first quarter phase, it is in its so-called *crescent phase;* when it is between its first quarter and last quarter phase, it is in its *gibbous phase* (from the Latin *gibbus,* meaning humped or swelled in size).

The moon's orbit is in a plane that is tilted about 5 degrees from the plane of the earth's orbit, which means that twice a year the moon is on a line with the sun and the earth. At these times, it is possible that the relative positions of the sun, earth, and moon will produce an *eclipse* (Fig. 7-16). The moon is sufficiently smaller than the earth to cause a lunar eclipse (the earth is between the sun and the moon) to black the moon out completely. A solar eclipse (the moon is between the sun and the earth) may completely black out the sun for only a few observers on the earth because only a small portion of the earth's surface will be in the moon's shadow.

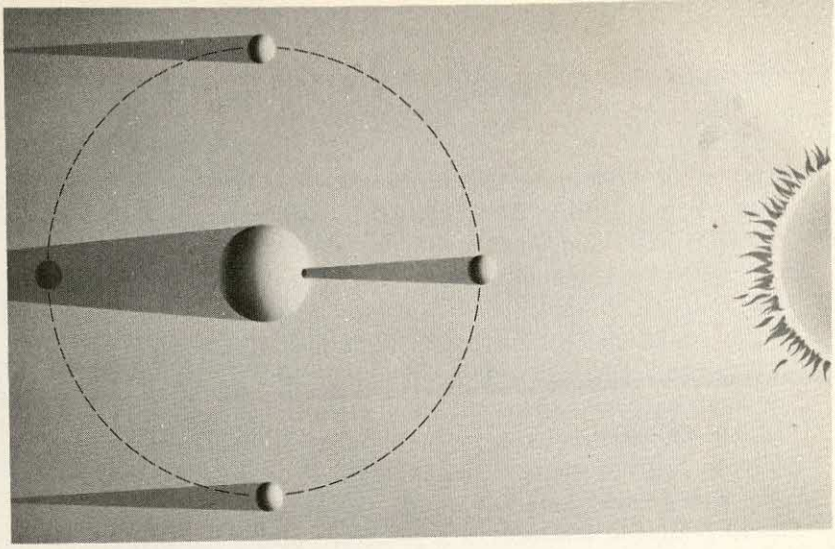


Fig. 7-16. Eclipses of the sun and the moon.

Our moon is one of the largest satellites in the solar system but other planets outdo us in quantity. Mars has two moons and Jupiter twelve, four of which are readily visible through ordinary field glasses. Saturn has nine satellites, Uranus five, and Neptune two. Mercury and Venus, however, have no satellites, and none have been discovered around the planet Pluto. Four of Jupiter's moons were first observed by Galileo with his primitive telescope. Since then, the moons of the planets have been one of the principal keys in determining the characteristics of their parents. For instance, the masses of the planets have been calculated by means of a rather complex relationship between the mass of the planet and the distance of the satellite from the planet—a relationship discovered by Newton. A moon's period of revolution depends upon the size of its orbit and the mass of the planet around which it travels. Hence, if we can measure the period of revolution of a moon and the size of its orbit, the mass of the planet can be calculated from the relationship:

$$\frac{\text{mass of planet}}{\text{mass of sun}} = \frac{(\text{satellite's distance from planet})^3}{(\text{satellite's period})^2} \times \frac{(\text{earth's period})^2}{(\text{earth's distance from sun})^3}$$

The accuracy of this method is as great as the entirely different method by which the mass of the earth was determined in Chapter 3. The diameter of a planet can be found by direct observation, and its mass and density can then be calculated (see Table 7-1).

7-10 Planetoids, Comets, and Meteors

In addition to the sun and the planets and their satellites, there are several other types of unusual and at times spectacular bodies that inhabit the solar system, many of which have been discovered only recently, indicating that more of them may be found as the methods of studying the sky are improved and refined. In 1772, Bode, the director of the Berlin observa-

tory, while examining the distances of the various planets from the sun (Table 7-1), noted that there was a huge gap between Mars and Jupiter. He thought that somewhere in this space there should logically be a planet, and his prediction touched off a search for it. Instead of one planet, astronomers found over a thousand small ones, revolving in almost a common orbit, 260,000,000 miles from the sun. The largest of these so-called *planetoids*, sometimes also called *asteroids*, is only 480 miles in diameter, and the smallest ones found so far are less than a mile in diameter. New ones are being discovered each year. Why there should be a multitude of small bodies in the orbit of a predicted planet remains one of the unsolved mysteries of astronomy.

Because we know more about *comets* now than the ancients did, we no longer believe that they are wayward objects bringing famine and pestilence in their wake. Soon after Kepler, many of them were found to be not only well-behaved but subject to the fundamental laws of planetary motions—that is, they followed elliptical orbits, the regularity of which makes it possible to predict their behavior. The most famous comet is named for Edmund Halley, an associate of Newton who first predicted that this comet can be seen every 75 years. Observed periodically since 87 B.C., it will next be visible in 1985. Its orbit is very elliptical, so that when it comes toward the sun, which is at one focus of its ellipse, it is greatly accelerated (Fig. 7-17). After a rapid swing around the sun, crossing inside and then outside the earth's orbit in only a few months, it passes on out beyond Pluto's orbit, where it slows down to less than a twelfth the velocity it has when near the sun. Halley's comet, therefore, spends most of its time well hidden in the outer reaches of the solar system.

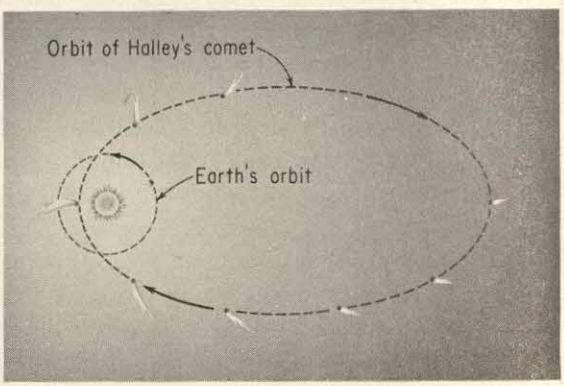


Fig. 7-17. The orbit and tail of Halley's comet.

All comets have a main body, which is bright and transparent, and some of them, such as Halley's have a luminous tail of dust or gas. As the comet comes near the sun, the radiation pressure of the sunlight drives smaller particles and gases from the comet's head and consequently, this tail swings away from the sun (Fig. 7-17). Comets are enormous, varying in sizes up to 100 times the size of the earth, and with tails as long as 50 million miles. Yet these objects have a density no greater than that of a typical laboratory vacuum. Some comets do not make regular appearances and may even leave our solar system. Some comets evidently disintegrate or scatter so that the pebbles, dust, and gas that compose them exert insufficient mutual attraction to hold themselves together. Thus, the broad expanse of space in the solar system is occupied by the numerous fragments of disintegrated comets, but, by comparison with our own atmosphere, these quantities are extremely small. The quantity of gas in the space between the planets, for instance, is considerably less than the amount of gas that remains in the ordinary vacuum that is produced in a laboratory. Only recently have vacuum pumps been designed that can approach the vacuum that exists in interplanetary space.

As the earth passes through the orbit of a disintegrated or "lost" comet, a shower of pebbles or chunks of stony material strike the earth's atmosphere. These fall rapidly through the atmosphere and are heated by friction with the atmosphere, causing brilliant streaks of light in the night sky, commonly called *meteors*. The earth encounters millions of meteors each day, but usually these visitors from outer space are burned up before they reach the earth's surface; when they do reach the ground they are called *meteorites*. The most famous meteorite of recent years is the one that flattened 750 square miles of forest in Siberia in 1908. In Arizona, the huge crater, measuring a mile in diameter and 600 feet in depth was probably blasted out by a meteorite about 50,000 years ago. Although the meteorite has never been found, thousands of small meteoritic pebbles are strewn around the crater, indicating that the main body of the meteor may have been vaporized.

Meteorites are all composed of substances that occur in the earth's crust. Some of them are stony, while others are metallic, containing mostly iron and nickle. They are direct proof that other parts of the solar system are made of the same kind of matter as the earth, and bring up the question of how the earth, the sun, and the planets were created. This perplexing question has baffled men for many centuries. To the many philosophic and religious explanations that have been proposed, physical science adds its explanation based upon observed data.

7-11 Theories on the Evolution of the Solar System

The theories on the origin and evolution of the earth and the solar system are only tentative and far from adequate. It is indeed a complex problem without sufficient evidence to solve it. That is, the information about the solar system, as derived by astronomers, geologists, chemists, space scientists, and others, points toward certain chains of events, but the gaps in the information are so big that it is most difficult to be certain about many of the steps in the development of the solar system.

One of the earliest recognized theories was proposed in 1755 by the German philosopher, Immanuel Kant. He proposed that it started from a huge primeval cloud that contracted and began rotating, resulting in the condensation of matter much the same way raindrops are produced in a cloud. The rotation caused a flattening of the mass, the sun at the center collected most of the matter, and the planets were formed in the outer parts of the region. This theory is still recognized today as one of the best, though, in view of the much greater amount of information now on hand, we are forced to view Kant's proposal as somewhat of a lucky guess.

The French astronomer and mathematician, Pierre-Simon

Laplace, in 1796 modified Kant's view by suggesting that the mother cloud, or *nebula,* was a rotating mass of low density whose original size exceeded that of the present solar system. As this spinning nebula cooled off, it shrank, and (according to Newton's third law) conserved its rotational momentum. That is, its speed of rotation increased as its size became smaller. This principle is illustrated by the twirling ice skater who rotates faster and faster as he draws his arms in toward his body and reduces the effective diameter of his rotating body. With our parent nebula, according to this theory, this increased rate of rotation caused rings to be thrown off that later condensed into planets. It is mathematically improbable that such rings would be thrown off, and since we now know of a ring of asteroids that failed to coalesce, it is doubtful if Laplace's rings would have formed planets.

In 1900, the geologist T. C. Chamberlain and the astronomer, F. R. Moulton, both of the University of Chicago, proposed that a collision of two stars resulted in a cluster of small particles that accumulated into planets as the hot masses shorn from the stars cooled down (Fig. 7-18). The collision theory was modified a few years later by Sir James Jeans and Harold Jeffreys of England by the suggestion that the two stars made a near approach in order to pull off small chunks of matter from each other. These theories have been largely discounted because of the incongruities of the effects of high temperature, light pressure, and the angular momentum of the various bodies in the solar system. For example, the idea that the earth was of high-temperature origin is not compatible with evidence found since these theories were in vogue.

The evidence that must be weighed in considering an adequate theory of the origin of the solar system includes several significant facts.

1. By far the largest part of the mass of the solar system is concentrated in the sun.
2. The planets, planetoids, and satellites, with only a few exceptions, revolve and rotate in the same direction in which the sun rotates.
3. The planets, planetoids, and satellites all revolve in orbits that are almost in the same plane as the sun's equator.
4. The sun rotates relatively slowly compared to the rotation rates of some of the planets.
5. The orbits of the planets and the planetoids are spaced at fairly regular intervals from the sun.
6. The members of the solar system are composed of similar materials, although the percentage composition varies. For instance, seven of the fifteen most common elements in the sun's atmosphere are among the fourteen most common elements on the earth.

7. The inner planets, Venus, Mars, Mercury, and particularly the Earth, seem to have been formed at relatively low temperatures, though they may have been completely molten at one stage in their development.

To explain these facts, several plausible theories have been proposed, none of which, however, has been fully accepted because we just do not have enough information to decide with any finality how the solar system was formed. Furthermore, it is exceedingly difficult to prove or disprove the validity of such a theory because of the very nature of the problem. In point of time and space it is humanly impossible to set up a laboratory experiment that would provide all the details needed for a conclusive answer.

One of the modern viewpoints, embodied in the ideas of C. F. von Weizäcker, the German astrophysicist, and others, including Harold C. Urey, the American astrochemist, reverts to a beginning of dust and gas clouds similar to those of Kant. At an early stage, these clouds condensed to produce many stars, of which the sun was one. The residual dust that was not part of the sun was gravitationally attracted to the sun and the gases were forced by light pressure away from the solar region. In a series of steps, involving turbulence, tidal effects of the sun, and accumulation of the dust into small bodies of the satellite and asteroid types, the planets were formed. The more dense ones developed near the sun, and the less dense ones were further out in the solar system. The planet bodies continued to grow by attracting to them the smaller bodies, and our moon appears to be one of these smaller bodies that was not captured. The asteroids as well as the moons of the other planets also remain as uncaptured objects in the solar system. The earth, incidentally, continues to capture many small objects of the meteor and meteorite class.

Of one thing we can be fairly certain: The earth's origin is closely related to the evolution of the solar system, and as we learn more and more of the processes by which the earth was formed, we will arrive at a better position to formulate a better theory of the evolution of the solar system.

7-12 Summary

The solar system, which has attracted man's curiosity for thousands of years, is best explained by a heliocentric theory in which the planets revolve in elliptical orbits around the sun. This, the Copernican theory, has supplanted the older Ptolemaic theory, which held that the earth is the center of the system and that the sun and the planets revolve about the earth in a very complex fashion.

A good scientific theory must fulfill certain criteria. A valid theory should explain a phenomenon in simple terms, permit and encourage the growth of new ideas, assist in making

valid predictions, and be capable of being modified and improved in order to fit new facts and ideas.

There are nine planets in the solar system and the inner four are much smaller than the outer five, with the exception of Pluto. In going from the inner to the outer planets, the period of revolution increases and the period of rotation decreases.

Kepler's three laws of planetary motion, which are proof that the solar system behaves predictably, describe the movement of the planets in elliptical orbits, the time interval during which a planet moves in its orbit, and the relationship between a planet's rate of revolution and its distance from the sun.

The sun, the center of the solar system, is a typical star and by far the largest member of the solar system.

The earth, with its single satellite, is a middle-sized planet and probably the only one where the temperatures are suitable for life as we know it. As the moon encircles the earth once a month, we see alternately more and less of its illuminated face, but only one face since the moon always presents the same side toward us. Solar and lunar eclipses occur whenever the earth and moon are in such a position that the sun's light to one or the other is cut off.

Planetoids, or small planets, occupy the region between the orbits of Mars and Jupiter. Comets, composed of dust or gas, may move in highly elliptical orbits in and out of the solar system. Meteors are stony or metallic objects that fall into the earth's atmosphere from outer space and are burned up by friction with the atmosphere. A meteor that reaches the earth's surface is called a meteorite.

The origin of the solar system is not well understood. Several theories have been proposed, but none of them is adequate because they do not explain all the facts now known.

EXERCISES

A. KEY TERMS TO REMEMBER

aberration of starlight
annual parallax
comet
crescent phase
ellipse
epicycle
equinox
geocentric theory
gibbous phase
heliocentric theory
Kepler's laws
lunar eclipse
meteor
meteorite
nebular theory
new moon
planet
planetesimal theory
planetoid
retrograde motion
revolution
rotation
satellite
solar eclipse
summer solstice

B. QUESTIONS ABOUT THE SOLAR SYSTEM

1. Do you think that the discovery of an annual parallax of the fixed stars at the time of Copernicus would have clinched his argument in favor of the heliocentric theory? Could the annual parallax be explained by more epicycles?

2. Evaluate or grade the Ptolemaic and Copernican theories on each of the requirements of a good scientific theory. Try to be impartial in your judgment, and disregard the eventual success of one of these theories.

3. On the basis of the information in Table 7-1, which of the planets should be most nearly a perfect sphere?

4. Describe briefly the information you would need to calculate the mass of a planet by means of observing its satellites.

5. When does Venus appear brighter—near the earth at crescent phase or farthest from the earth at full phase?

6. Why is the Antarctic polar icecap considerably larger than the Arctic polar icecap?

7. On the basis of the criteria of a good scientific theory and of the known facts, evaluate one of the theories of the origin of the solar system. Point out which facts are adequately explained and which ones are not.

C. PROBLEMS ABOUT THE SOLAR SYSTEM

1. Construct an ellipse with a small eccentricity and one with a noticeable eccentricity. What is an ellipse with maximum eccentricity?

2. From Table 7-1, select the distance from the sun and the period of revolution for two planets; substitute these values into Kepler's third law and see if the law is valid.

3. Examine Jupiter with ordinary field glasses, or a telescope, and observe the positions of its moons over a period of three or four hours.

4. Make a series of observations of the moon and record its changes of phase and time of rising.

5. Why are the warmest days in summer about a month after the summer solstice rather than near June 21. And why is the warmest hour of the day near 2 p.m. rather than at noon?

6. What spot on the earth's surface receives more light and heat from the sun than any other spot during a 24-hour period? Hint: This spot is not in the tropical zone.

SUGGESTED FURTHER READING

M. R. Cohen and I. E. Drabkin, *A Source Book in Greek Science* (New York: McGraw-Hill Book Company, 1948). Translations of the original Greek works on the movements of the sun, earth, moon, and planets.

Fred Hoyle, *Frontiers of Astronomy* (New York: Harper & Row, Publishers, 1955). A readable summary of recent observations of the solar system, with discussions of modern astronomical theories.

F. R. Moulton, *Consider the Heavens* (Garden City, N. Y.: Doubleday & Company, Inc., 1940). An easy-reading account of the development of astronomical science.

H. Shapley, S. Rapport, and H. Wright, *A Treasury of Science* (New York: Harper & Row, Publishers, 1946). An excellent

source book including several essays by Copernicus, Galileo, Moulton, Jeans, and Newton.

J. F. White, *Study of the Earth* (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1962). A collection of readings in geological science with several dealing with the evolution of the earth and the solar system.

G. deVaucouleurs, *Physics of the Planet Mars* (New York: The Macmillan Company, 1954).

Eight

Electrostatics and the Electron

In the last few chapters we have examined the basic laws of gravity, force, and motion, and we have found that these laws apply to many natural phenomena. The laws concerning mechanical energy apply to gross matter, that is, to relatively large quantities of matter compared to the minute particles of which matter itself is made. In order to explore the inner constitution of matter, where the classical laws of motion do not necessarily apply, we must acquaint ourselves with the laws of other kinds of energy—electrical and magnetic energy. Consequently, we now turn our attention from the gross properties of matter to its finer structure, and with this study of the characteristics of the smallest of particles we will be able to reach a more understandable picture of the physical world as a whole.

Ever since the time of the ancient Greeks, methods have been known for producing an electrical charge on an object. We can repeat some of the simplest of these methods, using materials close at hand. For instance, if you run a hard rubber or plastic comb through your hair and then place the comb near small pieces of paper, you will notice that it attracts the paper to it. When you scuff your shoes on a wool rug and then hold your finger near a radiator or water pipe, a noticeable spark will jump between your finger and the metal. Or, you may slide across the plastic seatcovers of your car and experience the snap of a spark as you touch the metal door handle. Another example, though less often observed, occurs when you quickly unroll a few inches of adhesive or friction tape in a dark room—at the point where the tape leaves the roll a bluish glow appears.

8-1 Electrostatics

All the phenomena mentioned above are manifestations of *electrostatics,* which means electricity at rest. A common method of producing an electrostatic charge is to stroke a hard rubber rod with a piece of cat fur or a wool cloth. The fact that the rod carries an electrical charge can be tested by bringing the rod near a suspended pith ball or a Ping-pong ball coated with aluminum paint (Fig. 8-1a). At first the ball jumps toward the rod, and then, after it touches the rod, it is repelled. The ball moves away even if chased with the rod (Fig. 8-1b).

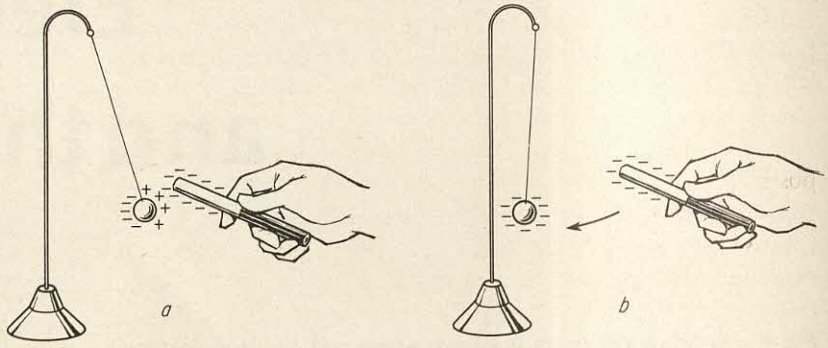


Fig. 8-1. The effect of a charged rod on a suspended metal-coated ball: (a) before contact and (b) after contact.

If a glass rod is rubbed with a silk cloth, on the other hand, and then brought near the ball in Fig. 8-1b, instead of being repelled, the ball is strongly attracted to the rod. However, as soon as the ball touches the glass rod once or twice, it loses its attractive force and the ball is repelled by the rod. But if we again bring in the rubber rod that has been rubbed with fur, we find that it again strongly attracts the suspended ball. Each time we change from a rubber rod to a glass rod we impart something different to the ball when we touch it that makes it strongly attractive to the other rod. This experiment, first performed several centuries ago, leads us to conclude that (1) the rubber rod picks up one kind of electrical charge from the fur and deposits this charge on the suspended ball, (2) the glass rod becomes oppositely charged when rubbed with the silk cloth, (3) these two kinds of electrical charge attract each other, (4) a rod may neutralize or change the charge on the ball, and (5) the same kind of charge on the rod and the ball results in repulsion. Our experiment, then, shows that there are two kinds of electrical charge, and that like charges repel each other but unlike charges attract each other.

In order to indicate the difference between these two electrical charges, we use the arbitrary designation made by Benjamin Franklin around the middle of the eighteenth century and say that the rubber rod is *negatively* charged after it is rubbed with fur and that the glass rod is *positively* charged after it is rubbed with silk. These electrical signs, in-

dicated in the accompanying figures, were adopted because of their convenience. They could as well have been called *A* and *B* or "he" and "she." Keep in mind that a negative charge does not mean a deficiency, such as we have when we overdraw our bank account. As we shall soon see, a negative charge, as well as a positive charge, is a quantitative attribute of matter.

An important point to note in all these experiments with electrostatics is that whenever an electrical charge is produced, an equal amount of the opposite charge is also produced. For instance, when the rubber rod is stroked with fur, not only does the rod become negatively charged but the fur becomes positively charged. This two-way action can be easily verified by holding the fur near the suspended ball after the ball has been touched by the rubber rod; the ball will act very much as it did when the charged glass rod was placed near it. That is, the fur bears the same kind of charge that the glass rod has. In the process of establishing a negative charge on the rubber rod, the fur becomes oppositely charged. This result indicates that when the two substances are rubbed together there is a transfer of electricity from one substance to the other, and it leads us to conclude that matter is composed of electrically charged particles that can be transferred from one point to another. In these experiments we are merely separating the negative charges from the positive charges by friction. Let us now see what constitutes these two kinds of electrical charges.

8-2 The Nature of Positive and Negative Electricity

Since the time of Benjamin Franklin, many scientists have studied the electrical properties of matter; yet even today, scientists do not know exactly what electricity is. We have learned much about how to measure and control it, but we have no good explanation why friction will cause electrostatic charges. However, after many successful, and unsuccessful, experiments, we now believe that electricity involves tiny particles of negatively charged matter called *electrons*. For instance, we think that at the beginning both the rod and the fur would have equal numbers of positive and negative charges on them; that is, they would be neutral. When the fur is rubbed over the rubber rod, the fur gives up some of its electrons to the rod which makes the rod negatively charged. At the same time, since the fur has lost some of its negative charge (which had just balanced the positive charges of the fur), it now has more positive charges than negative and appears as positively charged. There is evidently something in the constitution of the fur that permits us to remove electrons from it, leaving it positively charged. The nature of these positive charges will be defined in a later section. In summing up these ideas concerning electrical

charge, we find that:

positive charge = deficit of electrons
negative charge = excess of electrons
electric current = movement of electrons

The electron is thus the key to electricity. In order to get a clear picture of it, we shall devote a section later in this chapter to its discovery, its behavior, and its importance in the structure of matter.

8-3 The Electroscope

One of the first devices used to detect electrical charge was the *electroscope*. This instrument, shown in Fig. 8-2, consists of a metal rod, protected by glass, which supports two thin metal leaves at its bottom, one of which may be stiff or fixed. Either a positive or a negative charge placed on the metal rod will deflect the metal leaves, the amount of deflection being determined by the amount of charge on the rod. This deflection is caused by the force of repulsion between the leaves when they bear a like charge. If we touch the metal knob at the top with our hand, the leaves collapse because our hand neutralizes the charge on the electroscope and there is no longer an electrostatic force of repulsion to hold the leaves apart.

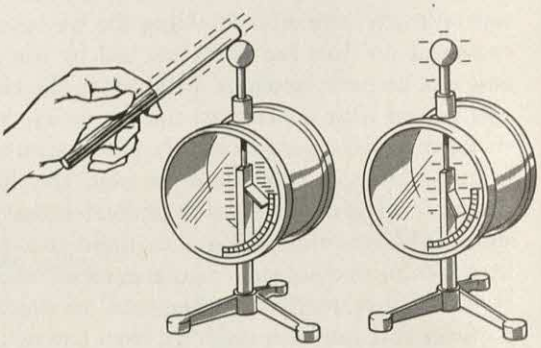


Fig. 8-2. An electroscope, showing how a negative charge can be stored on the instrument.

A very simple electroscope can be made by cutting a strip of newspaper from the top of the page to the bottom, allowing it to fold over at the middle and then placing the thumb outside, the index finger between the two strips, the middle finger at the other side, and stroking the paper downward. When a rubber rod that has been rubbed with fur is brought near the charged paper strips, there is a repulsion, thus indicating that the paper strips have a negative charge. With the paper electroscope, the charge on wood, glass, and metals can also be determined.

Two balloons floating at the ends of insulating threads may be used as an electroscope as shown in Fig. 8-3. Charges can move freely over the balloon's surfaces if we coat them with a light film of graphite. When the balloons are neutral as in (*A*), they hang straight down with only a very slight gravitational attraction pulling them together. If the balloons are charged oppositely as in (*B*), a marked attraction occurs, while a repulsion takes place when their charges are alike.

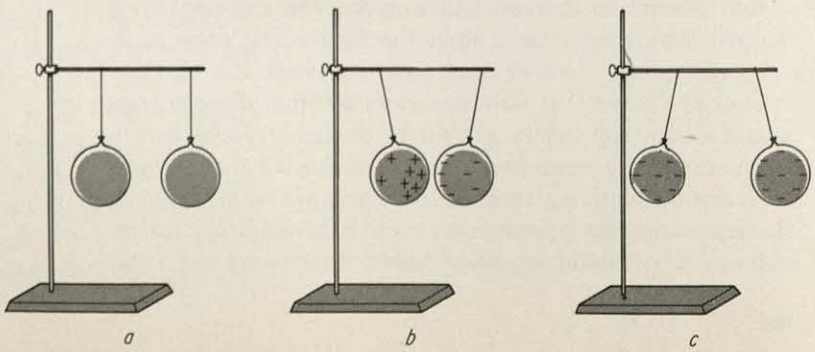


Fig. 8-3. Balloons hanging by a thread from a wooden rod show attractive force between unlike charges and repulsion between like charges.

8-4 Insulators and Conductors

Materials may be placed in one of two general classes depending on their ability to hold and carry an electrical charge. For instance, a copper ball mounted on a glass rod can be negatively charged by rubbing it with fur, if we do not touch it with our hand during the process, for the moment the ball is touched it loses its charge. On the other hand, a rubber or glass ball can be charged by friction and not lose its charge when we touch it with our hand. This behavior indicates that there is a fundamental difference between the one class of materials—metals—and the other class of materials—rubber, glass, porcelain, plastics, sulfur, and so forth. We call the metals *conductors,* because a charge moves along them very rapidly, and the other materials *insulators,* because they tend to retain their charge. No amount of rubbing with either fur or silk will impart a charge to a metal rod held in our hand, because the moment any charge is formed it is drained off by our hand, which is a fair conductor itself and which in this instance is called a *ground* because it has great capacity to hold both positive and negative charges. Since the earth is the most satisfactory ground, if we wish to dispose of an electrical charge, we simply drive a metal rod into the ground, and all the electrical charge drains off into it.

A simple experiment that demonstrates the difference between conductors and insulators is shown in Fig. 8-4. When a charged rubber rod is touched to a metal ball *A*, which is supported by an insulator, the charge travels along the copper wire *B* to the electroscope, where it deflects the leaf. This experiment shows that a charge can move through a conductor to the electroscope, but not through the insulator that supports the metal ball. The difference between these two classes of materials is essentially this: In metals, the electrons are free to move, but in insulators the electrons are tightly bound. When electrons enter a metal bar they repel the electrons in the metal immediately adjacent to them, and these electrons in turn displace the electrons next to them and so on in a progressing reaction throughout the bar. Thus, in Fig. 8-4, the electrons that deflect the leaf of the electroscope are not necessarily the same ones that were on the charged rubber rod.

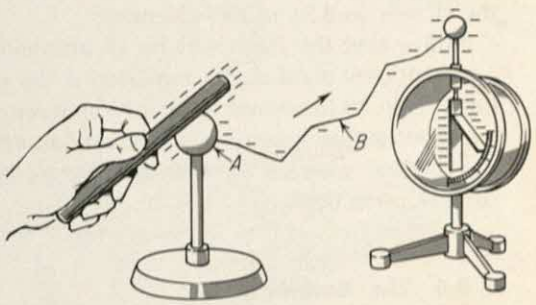


Fig. 8-4. An experiment showing that a metal is a conductor of electrical charge.

8-5 Electrostatic Attraction and Repulsion: Coulomb's Law

We have found that like charges repel and unlike charges attract each other. The strength of these repulsion and attraction forces was first measured in 1780 by Charles A. Coulomb, who found that *the force between two electrostatic charges is directly proportional to the product of the magnitudes of their two charges and inversely proportional to the*

square of the distance between them. This relationship sounds much like the law of gravity, and, when expressed as a mathematical equation, is indeed similar to Newton's gravitational law:

$$F = k\frac{q_1 q_2}{r^2} \tag{8.1}$$

Here, F is the force between the two charges, q_1 and q_2 are the charges on the two objects, and r is the distance between them. The value of k, which is a constant of proportionality, will depend upon the units we choose for the other quantities. If the charge is in coulombs, the distance in meters, and the force is in newtons, k will be approximately 9×10^9 $\frac{\text{newton m}^2}{\text{coulomb}^2}$.

A coulomb of charge could be defined as the charge which when placed 1 m away from a like charge will repel that charge with a force of 1 newton when the constant k has the above value. The coulomb, while actually defined on the basis of the force between moving charges, is equivalent to the charge on 6.24×10^{18} electrons.

Notice that the force will be an attraction if the charges have opposite signs and a repulsion if the charges have the same sign. An important difference between gravity and electrostatics is that gravity provides only an attraction between two bodies, whereas electrostatic charges can either attract or repel each other.

8-6 The Electron

Since we have found in our consideration of electrostatics that the electron is the key to electricity, let us now see how it was discovered in order that we might become better acquainted with the electrical nature of matter. Few scientific pursuits have been more productive than that which revealed the existence of electrons. No one has ever seen these tiny particles, yet there is almost overwhelming evidence that they do exist.

In 1853, Masson, a Frenchman, devised an experiment in which he passed a spark through a glass tube from which most of the air had been removed. The spark, as it traveled through a partial vacuum, appeared as a soft glow, and it followed a wide path and filled the whole tube with a distinctive color. In contrast to a spark passing through a partial vacuum, a spark that passes through air is bright and intense and follows a sharply defined path. The neon advertising signs that line our main streets are the practical outcome of Masson's simple experiment. As better vacuum pumps were built, many scientists studied the conduction of an electrical charge through gases at low pressures. One of these pioneers was Sir William Crookes of England, who in 1870 developed the famous Crookes tube, or cathode ray

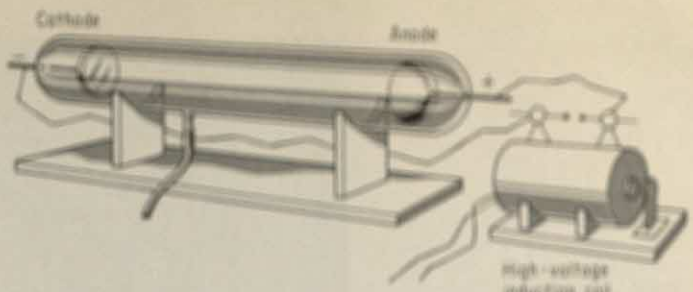


Fig. 8-5. A simple cathode ray tube.

tube, which was the crude ancestor of the modern television picture tube. To many workers goes the credit for studying cathode rays, but it was J. J. Thomson in 1897 who first explained their true nature.

The simplest cathode ray tube is a glass tube, with metal disks sealed into each end, attached to a vacuum pump (Fig. 8-5). When an electrical current is passed through the tube, long streaming sparks travel from one end of it to the other. As the vacuum pump reduces the amount of air in the tube, the spark turns into a faint purplish column, which becomes brighter but then is eventually broken into striations or thin bands. The dark gaps grow until the tube becomes almost colorless, and finally, at a very low pressure, the tube takes on a soft greenish glow which spreads and then becomes quite bright. The green glow is attributed to what we call *cathode rays*, which, although invisible, cause a number of phenomena that are readily observable. The properties of these rays, studied by Crookes in 1879, are most significant for modern physical science:

1. They travel in straight lines and cause shadows to be cast by objects placed in their path (Fig 8-6a).

2. They are capable of producing mechanical motion, for they can turn a paddle wheel inside an evacuated tube (Fig. 8-6b).

3. They produce fluorescence in many objects upon which they fall in the form of a soft greenish glow similar to the light emitted by the luminous numerals on a watch.

4. They produce a rise in temperature in the objects they strike, that is, the walls of a Crookes tube actually become warmer, showing that these rays have energy which is converted into heat as they strike the walls of the tube.

5. They may be deflected from straight line paths by a magnetic field (Fig. 8-7). This is a simple experiment that may be performed by anyone with an ordinary horseshoe magnet as shown in the figure, and it shows that there is a relationship between the rays and the magnet's effect.

6. The charge carried by the cathode rays is negative, since the rays are attracted to a positively charged plate that is placed along their path.

7. They can penetrate thin sheets of metal. Since the energy of the extremely small particles that make up cathode rays varies, the depth of their penetration varies, but ordinarily, they can pass through aluminum 1 mm thick.

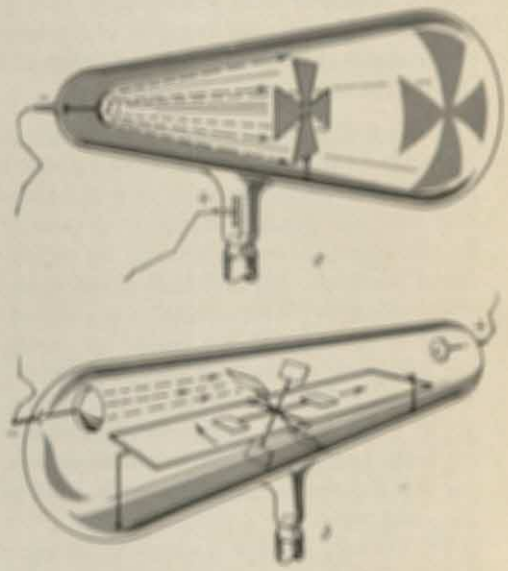
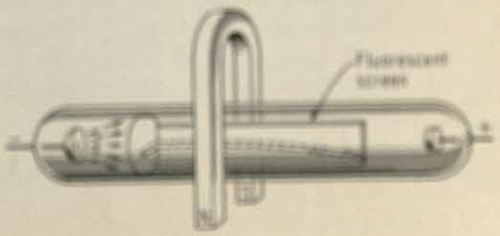


Fig. 8-6. Two characteristics of cathode rays: (a) they travel in straight lines and (b) they produce mechanical energy.

Fig. 8-7. The deflection of cathode rays by a magnetic field.



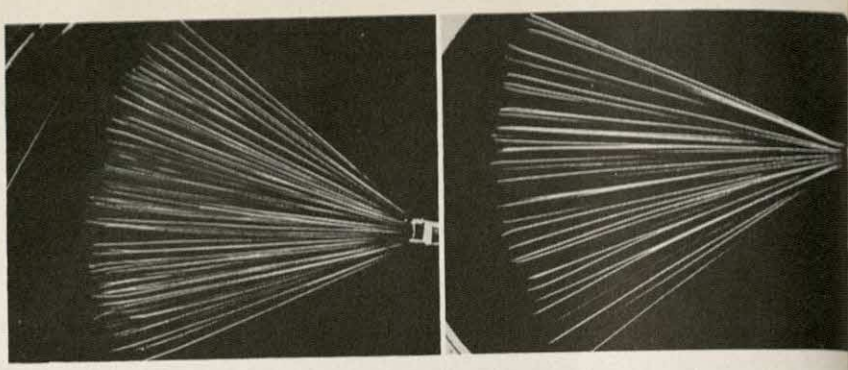


Fig. 8-8. Fog tracks made by alpha particles emitted by a radioactive substance. (From Gentnev, Maier-Leibuitz, and Bothe, An Atlas of Typical Expansion Chamber Photographs. London: Pergamon Press Ltd., 1954)

8. They act as nuclei for the condensation of supersaturated vapors, thus producing fog tracks similar to those in Fig. 8-8. That is, if we pass a cathode ray through a region that is saturated with water vapor (100 per cent relative humidity), the rays act as points of condensation so that the water vapor condenses into fine droplets of fog and thus indicates the actual path of the rays.

It was then established by Thomson that these cathode rays are simply a stream of electrons that passes from the cathode through the rarefied gas to the opposite end of the tube. Since that time, the negative charge on a cathode ray particle, or an electron, has been determined by no less than eleven different methods, all of which agree upon the amount of charge it carries. Since no smaller quantity of electrical charge has ever been found, the charge on an electron is used as the basic unit of negative charge. The electron is approximately 5×10^{-13} cm in diameter, and its mass, of 9.1072×10^{-31} kg, is only $\frac{1}{1836}$ the mass of the hydrogen atom. Since the hydrogen atom is the smallest atom there is, we can see that the electron is an extremely small particle in charge, size, and mass.

With this knowledge of electrons, we can now postulate an explanation for our electrostatics experiments. Since the charge of electrons and the charge of a negatively charged rubber rod are similar, we can conclude that the rubber rod bears an excess of electrons on its surface. These electrons are stripped from the surface of the fur during the rubbing process, leaving the fur positively charged. If the positive charge on the fur is the result of removing electrons from the fur, there must be particles larger than electrons in the fur that contain both positive and negative charges, since the removal of some of the electrons leaves a positive charge on the fur. All of this indicates that the fur must be composed of particles, or *atoms*, much larger than an electron, that contain both electrons and some kind of positive charges.

At almost the same time that Thomson explained cathode rays and electrons, a group of scientists discovered a positively charged particle that came from inside the atom. This discovery resulted from the study of the radioactivity of certain substances and provided a powerful tool with which to probe the inner constitution of matter itself. As is true with

many discoveries in science, the existence of a positive charge within an atom was found by a series of experiments that were aimed at an entirely different objective.

8-7 Radioactivity

Radioactivity was discovered by accident. In 1896, Henri Becquerel of France was trying to see whether there was a connection between the recently discovered X rays and fluorescence or phosphorescence. To do this he would prepare a photographic plate (they had to sensitize their own film in those days), then wrap it in dark paper, place a salt of uranium (or other fluorescent material) on top of the dark paper, and expose the salt to the sunlight in order that it fluoresce strongly. He successfully found that the fluorescing salt seemed to provide X rays. One day when he had the film prepared and the uranium on top of the paper covering, it became cloudy and he put the whole thing into his desk drawer. After several cloudy days he developed the plate, expecting to find the image very dim, but on the contrary, the image appeared "with great intensity." The effect he had at first observed did not depend upon the fluorescence or phophoresence but rather on some other activity of the uranium. This is now called *radioactivity*.

Intrigued by Becquerel's findings, a number of scientists turned to the study of this peculiar emanation. Among them were Pierre and Marie Curie of France, who succeeded in isolating new elements—including polonium and radium—that were even more powerful than uranium in producing these invisible, penetrating rays.

The emanations given off by a radioactive substance were soon found to possess unusual properties. When a radioactive substance was placed in an atmosphere supersaturated with water vapor, the rays produced fog tracks (Fig. 8-8). They also caused fluorescence, or glowing, in such substances as zinc sulfide and other minerals, and they were found to have electrical properties (Fig. 8-9). The behavior of the emanations from a radioactive material indicated that they were composed of three types of rays, α (*alpha*), β (*beta*), and γ (*gamma*), the properties of which are summarized in Table 8-1.

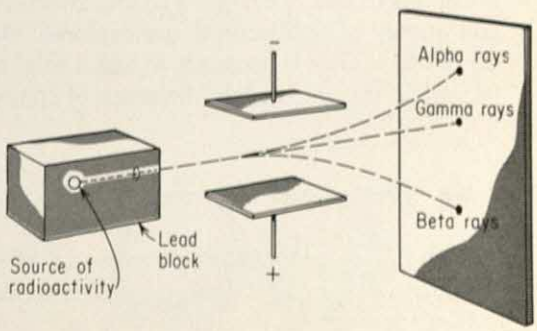


Fig. 8-9. The effect of an electrical field upon radioactive rays.

The alpha ray, composed of positively charged particles, is evidence that there are positive charges inside the atom in addition to the negative charges that are manifested in both electrostatics and in beta rays. Thus, we have experimental proof that the atom contains both positive and negative charges. It is important to note, however, that the negatively charged electrons can be removed quite easily by friction, as in our electrostatics experiments, whereas the positive charges can be removed only by much more elaborate processes such as radioactivity. We shall refer again to radioactivity, espe-

TABLE 8-1 PROPERTIES OF FUNDAMENTAL PARTICLES AND RADIOACTIVE EMANATIONS

Name	*Nature and Make-up*
Electron	1. Particle 2. Mass $= 9.1 \times 10^{-31}$ kg $= \frac{1}{1836}$ unit * 3. Charge = negative $= 1.6 \times 10^{-19}$ coulombs 4. Located in outer part of atoms
Proton	1. Particle 2. Mass $= 1.67 \times 10^{-27}$ kg = 1 unit 3. Charge = positive $= 1.6 \times 10^{-19}$ coulombs 4. Located in inner part of atoms
Neutron	1. Particle 2. Mass $= 1.67 \times 10^{-27}$ kg = 1 unit 3. Charge = none 4. Located in inner part of atoms 5. Free neutron very penetrating
Beta rays	1. Identical with electron 2. Penetrates up to 2 m of air
Alpha rays	1. Particle 2. Mass $= 6.68 \times 10^{-27}$ kg 3. Charge = positive $= 3.2 \times 10^{-19}$ coulombs 4. Consists of 2 protons and 2 neutrons 5. Low penetrating power (5 cm of air) 6. Induces electrical charges on other particles
Gamma rays	1. Electromagnetic wave (similar to X rays) 2. No mass or charge 3. Very penetrating

* Mass units are based on the mass of the hydrogen atom as unity. Thus, the alpha particle has four times the mass of a hydrogen atom.

cially in Chapter 12 where the detailed structure of the atom is examined, and in Chapters 25 and 26 where the mechanism and energy of radioactivity are explored. Our present interest in this subject is primarily to find a solid basis upon which to explain the electrostatic behavior of substances.

8-8 Protons and Neutrons

With this new tool—radioactivity—scientists began a series of experiments in which these emanations were used as projectiles for studying the inner composition of matter. A beta particle, for instance, can travel through 7000 atoms of gold

before it is deflected from its original path, and an alpha particle can pass through as many as 15,000 atoms of gold before it undergoes an appreciable deflection. These experiments confirm a conclusion that was reached early in the study of the structure of matter, namely, that atoms of matter are mostly space and that the ultimate particles composing matter are extremely small compared to the atoms themselves. These experiments also proved that matter contains positively charged particles, one of which is the alpha particle that possesses two unit positive charges and a mass several times that of the smallest atom. Hence, the single particle in matter that is positively charged must be smaller than an alpha particle. We call this single, positively charged particle the *proton*, and its mass is essentially the same as the hydrogen atom. The mass and charge of protons were first measured by Lord Rutherford of England who bombarded nitrogen with alpha particles.

For some time, scientists believed that electrons and protons were the only particles that made up an atom. In 1932, however, Chadwick, an Englishman, discovered the existence of a neutral particle—called the *neutron*—that is similar in mass to the proton. Because the neutron carries no charge it obviously could not be detected or even suggested by electrostatic experiments, and it remained undetected until more powerful methods were found for studying the inner parts of the atom.

The process of predicting before discovering is typical of the way science sometimes works. Even though the neutron was not discovered until some twenty or thirty years after the electron and proton were established, scientists long suspected that a neutral particle existed in the atom. For instance, with the alpha particle's mass of four units and two positive charges, there remained the problem of accounting for the extra two units of mass, which led scientists to believe that a neutral particle must be part of an alpha particle. This, briefly, is a good example of the way science works—a dependable theory, based on sound experimental evidence, points the way for the scientist and often guides him to the answer of a problem before the answer is proved experimentally.

8-9 Induction

It is possible to build up a charge on a body without touching it by a method known as *induction*, illustrated by the experiment in Fig. 8-10. Two metal spheres, *A* and *B*, sup-

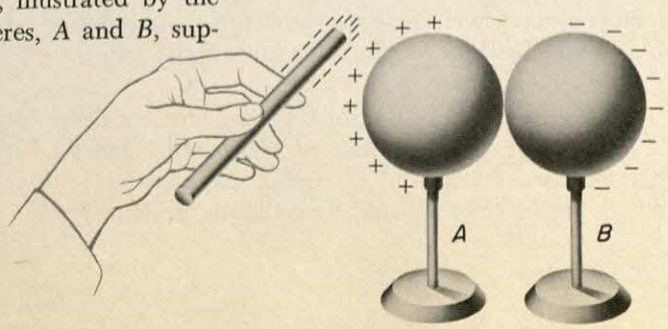


Fig. 8-10. The charging of a metal ball by induction.

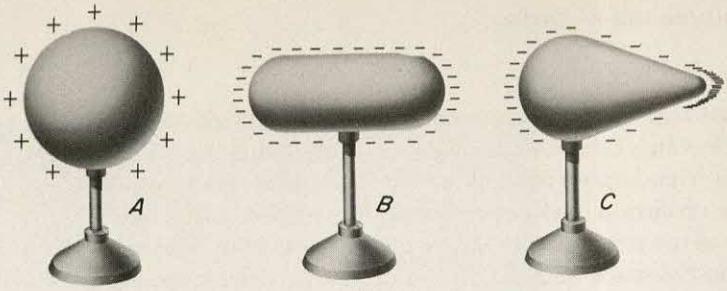


Fig. 8-11. The distribution of electrical charge on metallic objects of various shapes.

ported by glass insulators, are touching each other when a negatively charged rod is brought near *A*. No charge is actually transferred from the rod to sphere *A*, but after the spheres are separated and the rod removed from their vicinity, each sphere will be charged, *A* positively and *B* negatively. The charges now carried by each sphere are called induced charges. The explanation for these induced charges is based on your understanding of electrons. When the negatively charged rod is brought near the pair of spheres, electrons are repelled by the charge on the rod from sphere *A* to sphere *B*, leaving *A* with a deficiency and *B* with an excess of electrons. When the spheres are parted, each one retains its unbalance of electrons, or its charge.

Electrical charges on insulators remain where they were originally deposited, but on conductors the charge distributes itself over the entire surface of the object. On a spherical conductor, such as *A* in Fig. 8-11, the charge is distributed uniformly over the surface. On irregularly shaped conductors, however, the charge is distributed according to the shape of the surface, with the most concentrated charge at the points of maximum curvature, as shown on *B* and *C* in Fig. 8-11.

Many interesting experiments can be conducted to illustrate electrical induction. A classic example is the famous Kelvin "water dropper," shown in Fig. 8-12. If a stream of water is divided into two equal parts and allowed to drip through metal tubes, a charge is built up sufficient to produce a noticeable spark at *A*. The spark indicates that so great a charge has been built up on *B* and *C* that the air, ordinarily an insulator, conducts the charge between the two cans and thus neutralizes the induced charge. Drops of moving water pick up a charge, much as a rubber rod picks up a charge from fur. From the diagram, note that a slight imbalance in charge between *D* and *E*, caused merely by chance, gives the droplets of water at *D* a negative charge, those at *E* a positive charge, and these charged droplets in turn induce a charge of the opposite sign on the metal tubes through which they fall. Large food cans, with their tops and bottoms removed, work well as these metal tubes. As much as 7000 volts of electrical potential can be built up by induction on the two sets of cans.

A very practical application of induction is made in a device called a *capacitor*, which consists of two parallel metal plates separated by an insulator. If one of these plates is charged either by an electrostatic charge or by a battery,

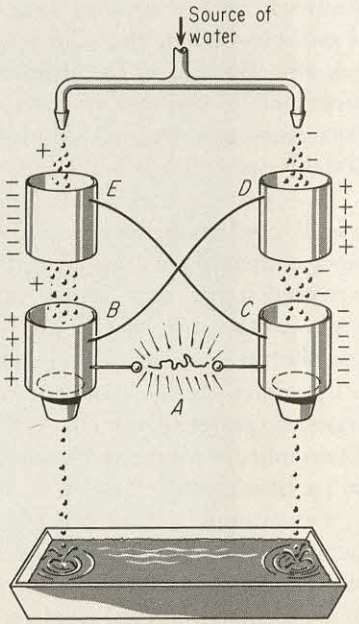


Fig. 8-12. Kelvin's "water-dropper" generator of electrical charge by induction.

electrons cannot flow across the insulator, but an opposite charge will be built up on the second metal plate as the result of induction. A capacitor is often used when an oscillating charge—which makes first one plate positive, then negative, and so on—is placed on it, and such a capacitor is extremely useful in the curcuits employed in telephones, radios, television sets, and other electronic appliances.

8-10 Electrostatics in Nature: Lightning and Aurora Borealis

In the laboratory, scientists have manufactured lightning that leaps as far as 60 feet through the air. During an electrical storm, however, nature produces a magnificent show of electrostatics that dwarfs man's attempts to produce lightning. Nature's method of making electrostatic charges differs from our electrostatic experiments of rubbing a rod with fur, for her lightning results from the separation of charges within a rain cloud. A storm cloud is a turbulent region in which rain droplets form, fall a short way, and then may be carried upward by rising winds; the droplets may pass up and down several times before they fall to the earth. During this time the droplets grow, but after they reach a certain size the violent air currents usually tear them apart, with the smaller drops being carried upward more rapidly than the larger ones. It is during this separation process that a charge is formed, according to the principles of electrostatics that we have already examined.

Electrons, even in a conductor, are more concentrated on the surface than in the body of the material (Fig. 8-11), so water, which is only a fair conductor, accumulates more negative charge on its surface than inside. Since the outer part of a raindrop is negatively charged and the inner part positively charged, the small fragments that are broken off by the rapid updrafts are negatively charged (Fig. 8-13).

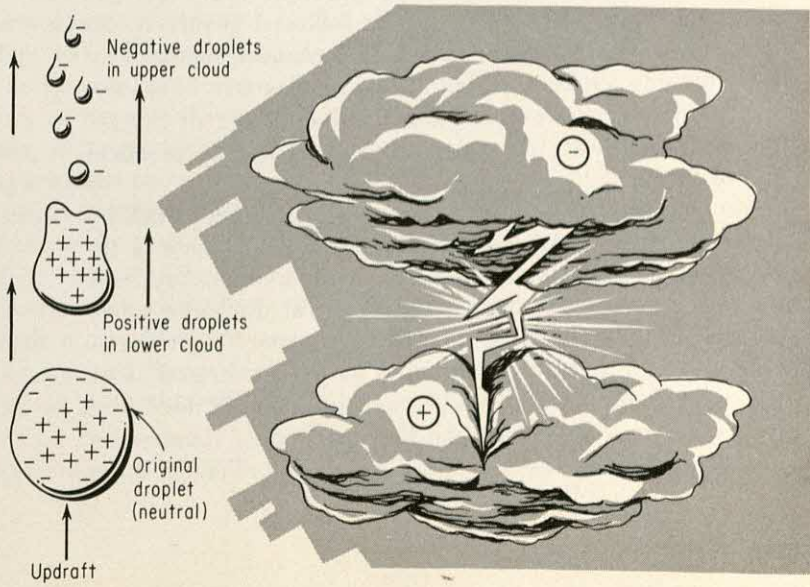


Fig. 8-13. One explanation of the formation of electrical charges on storm clouds.

These negative droplets form the upper part of the cloud and the larger positively charged droplets form the lower part of the cloud. Eventually the two charges may become so great that they have to neutralize themselves by a lightning discharge between the clouds, or occasionally between the cloud and the earth, producing a gigantic spark that may be several miles long.

Other scientists think that some of the water droplets must be frozen before electrical charges will separate in large quantities. They feel that as the water droplet is swept high enough that ice will form; the ice in forming may pick up more liquid water than can be immediately frozen and some of the water will be torn away by the air and carried upward. Under these circumstances, it is thought that the water droplets would have a positive charge and that the upper part of a thunder cloud would then be positive. This is found to be the case in many storms and often there is also a positive portion at the cloud base, so quite possibly both processes, together with still others at present unknown, are at work.

The production of electricity in a cloud and thunderstorm is not very well understood and we have presented only two theories. Regardless of whether the processes described here are the actual ones active in a particular storm, the principles of induction and frictional electricity which you have been studying are applicable to this gigantic display of nature's force.

The lightning stroke itself is interesting. A surge of charge moves downward from the cloud for about 150 feet, taking 50 μsec for the movement, pauses an instant, then moves again in a series of similar jerky downward surges. It takes between 1/1000 and 1/100 sec for this charge to reach the earth. This electrical surge does not produce enough light for it to be observed ordinarily, but as it reaches the ground, another surge of charges moves up this prepared path at about 10^7 m/sec. This carries an average current of 10,000 amperes and produces enough light and heat to be seen and to produce the thunderclaps of sound. After 1/100 sec, another stroke may take place and be followed in turn by 30 or so strokes in the same channel. This channel, followed by the lightning stroke, is about 10 cm in diameter. It is estimated that there are about 16,000,000 lightning discharges on earth each year, but so far man has made little practical use of them.

Skyscrapers, tall buildings, and smokestacks are struck by lightning many times every year, but little damage is inflicted upon them because most of these structures are built on steel frames that conduct the electrical discharge safely to the ground. The thunder that accompanies lightning is caused by the sudden expansion of air as it is heated along the path of the lightning spark, and the sound is intensified by echoes from the clouds and the surrounding countryside.

Another dramatic electrostatic phenomenon of nature is the

Fig. 8-14. Aurora borealis as observed near the Arctic Circle. (Courtesy C. T. Elvey, University of Alaska)

peculiar curtainlike glow called *aurora borealis,* or "Northern Lights," in the Northern Hemisphere (Fig. 8-14). The aurora occurs most prominently at the North and South poles, but at times it may spread over other regions of the earth. One of the better explanations for the aurora, although there is considerable difference of opinion among scientists about this, is that charged particles from the sun cause the particles of the upper atmosphere, 60 to 600 miles above the surface of the earth, to become charged and, possibly influenced by the earth's magnetism, to emit a green-yellow or red glow. By comparing observations of the aurora made at two locations, scientists have set the height of the aurora, and therefore the outer limit of the earth's atmosphere, at upward of 600 miles.

Not nearly so exciting, but having growing importance to

our smog-plagued cities is the smoke precipitator. This device uses the electrical charges on smoke particles to attract them to charged plates where they are electrically discharged and are removed from the flue gases. This process is described in Chapter 21, Fig. 21-14.

8-11 Summary

Electrostatics supplies us with an introduction to the electrical nature of matter. There are two kinds of electrical charge—positive and negative—and they are caused by a deficit and an excess of electrons, respectively. These charges can be readily detected by using an electroscope.

Electrons are unit negatively charged particles with a mass $\frac{1}{1836}$ of the hydrogen atom. When electrons are emitted in a vacuum tube they form cathode rays, the properties of which reveal the characteristics of the electrons themselves.

Materials are classed as insulators or conductors depending on their ability to hold or transmit an electrostatic charge. Metals are good conductors and readily transmit electrons from one point to another, while glass, wood, and rubber are examples of insulators that do not transmit electrons but are capable of storing up electrical charge on their surfaces.

Like electrical charges repel each other, unlike charges attract. Electrical charges obey Coulomb's law, which says that charges exert a force on each other that is proportional to the product of the charge and inversely proportional to the square of the distance between them.

Radioactivity gives us a key to the inner structure of matter. The radioactive emanations—alpha rays, beta rays, and gamma rays—are useful in studying the structure of matter, for they penetrate matter easily, indicating that matter is mostly space.

The unit positive particle is a proton, which has a mass about the same as the hydrogen atom. The counterpart of the proton is the neutron, a particle that has about the same mass as the proton but no charge.

An electrical charge on one object can induce an opposite electrical charge on a neighboring object in a process called induction. In nature, a bolt of lightning is an enormous spark that streaks between a positively charged cloud and the negatively charged ground or between two clouds, one positively charged and the other negatively charged.

EXERCISES

A. KEY TERMS TO REMEMBER

alpha ray	electron	lightning
aurora borealis	electroscope	negative charge

beta ray	electrostatics	neutron
cathode rays	electrostatic unit	positive charge
capacitor	gamma ray	proton
conductor	induction	radioactivity
Coulomb's law	insulator	

B. QUESTIONS ABOUT ELECTROSTATICS AND THE ELECTRON

1. Devise a means to test the sign of the charge on the surface of a piece of varnished wood, a glass window pane, a piece of clothing, or a sheet of paper.

2. What would happen if a lighted match were brought near the knob of a charged electroscope? Explain in terms of electrons.

3. Why is it impossible to produce a negatively charged object without producing a positively charged object at the same time?

4. What is a safe location during an electrical storm? What are several unsafe locations? Why are these latter locations unsafe?

5. Why are lightning rods placed on the highest points of a building? Why are the rods pointed at their upper end?

6. If matter is made up largely of electrical charges, why are not common substances such as wood, glass, and iron either positively or negatively charged?

7. Describe an experiment which would show that there are two kinds of electrical charge.

8. If you had a negatively charged insulator and also an uncharged metal sphere on an insulating stand, how could you put a positive charge on the metal sphere?

9. How can you charge an electroscope either positively or negatively from a positively charged glass rod?

10. Why is attraction not a good test of the kind of electrical charge an object has?

C. PROBLEMS ABOUT ELECTROSTATICS AND THE ELECTRON

1. When the static charge in Coulomb's law is in coulombs and the distance is in meters, the force is in newtons, calculate the force that exists between two positive charges of 0.5 coulombs each that are 20 cm apart.

2. What would be the force of repulsion in newtons between two electrons 1 cm apart? Hint: The answer will be extremely small.

3. Make a graph plotting the distance from an electrical charge versus the force acting on a like charge.

4. A +100 microcoulomb charge is placed at the 0-cm mark on a meter stick. If a second +100 microcoulomb charge is placed at the 30-cm mark, how much force would these two charges exert on a third +100 microcoulomb charge placed at the 40-cm mark?

5. Two small pith balls are suspended one above the other at a distance of 0.04 meters. If they each have a mass of 2×10^{-4} kg and the lower one has a charge of $+4 \times 10^{-12}$ coulombs, what must be the charge on the upper ball in order to lift the lower one?

6. An alpha particle and an electron are 1 cm apart. Which is the greater, the gravitational or electrical attraction between them?

Nine

Electricity and Magnetism

Legend tells us that the ancient shepherd, Magnes, found it difficult to climb down from Mt. Ida near ancient Troy because the tip of his staff and the nails in his shoes were strongly attracted to the ground. Ulysses, in Homer's *Iliad*, suffered a more disastrous fate—his ship was drawn by a mysterious force toward a large black rock upon which he was subsequently shipwrecked. Although centuries later scientists did find a naturally occurring substance, called lodestone, that attracts metallic objects, its force is so weak that it takes a sensitive instrument to detect it, and we must, therefore, regretfully conclude that these legends are pure fiction. Before the twelfth century, however, it was discovered that a lodestone, when suspended by a string, orients itself in a special way in a north-south direction. This discovery led to the invention of the compass, which became an indispensable guide for the mariner who pushed out from Europe's shores to explore the New World.

Today, however, powerful magnets have been put to work for us, magnets that we have produced ourselves. They enable us to generate electricity, which in turn runs our radios, telephones, television sets, electric motors and lights, and numerous other appliances. In order to understand how these conveniences are powered, let us first examine our two invaluable servants—electricity and magnetism—separately, and then in harness together. In the process, we will also find out a great deal about the basic structure of our physical world.

9-1 Electricity

Our modern concept of electricity stems from the famous experiments conducted by Benjamin Franklin around the middle of the eighteenth century. From his kite experiment, Franklin reasoned that electricity consists of a substance, the presence of which produces a positive charge and the absence of which produces a negative charge. Electric current, he believed, is the flow of this substance from a point where it is in excess to a point where it is lacking; when there is neither an excess nor a deficiency, there is no current. To the excess or deficient condition he assigned the corresponding + and − signs with which we are all familiar. The two terminals of a battery are labeled + and −, as are parts of many other electrical devices that must be attached to either one or the other of these two poles. Franklin never went so far as to attribute any particular properties or characteristics to the substance he called electricity. Only after many years of study and experimentation have we been able to determine that it is a process involving the movement of charged particles called electrons. This movement of electrons is responsible for several important effects, such as heat, light, and magnetism, but before we examine these effects, let us become better acquainted with the movement of these tiny particles.

In the last chapter, we saw how static charges can be built up on hard rubber, glass, and other insulators. Unlike static charges can be neutralized, however, if we connect the charges together by a conductor such as a copper wire. In terms of electrons, the negatively charged rubber rod is overloaded with electrons so the excess ones tend to escape through the conductor to the glass rod which is hungry for electrons. By constantly rubbing glass and rubber rods, we could generate an electric current, but this method is not very efficient because the amount of electricity produced is extremely small. It would take, for instance, at least five million men rubbing continuously to produce enough electricity for a typical American family.

Fortunately for us, there are other ways of generating electricity that are much more efficient. One is the method used in batteries, in which an exchange of electrons is created between two poles by means of a chemical reaction. The ordinary flashlight battery operates in this way, and we shall refer to this method in Chapter 19 on chemical processes. Even if all the batteries manufactured in the United States during a year, however, were used to light New York City, they would last for only one night. Another method, and the one that gives us most of our power, is produced by electromagnetic effects. In order to understand this method, we need to know more about both electricity and magnetism and about the units used to measure them.

The effect of electricity travels almost at the speed of light. Our natural reaction to this information is to ask: How is it able to go so fast? Metallic conductors, such as silver,

copper, and iron, are composed of atoms, each of which has one or two electrons that are free to move from atom to atom. When an electric current is passed through a conductor, its speed is determined by the ease with which electrons move from one atom to the next. In Fig. 9-1 only those electrons that are free to move are shown. To simplify matters, consider that each atom possesses one of these free electrons at all times and that they can be easily exchanged with other atoms. If a single electron enters one end of the conductor in Fig. 9-1 at *A*, it will cause each electron to move ahead slightly and thereby force an electron out the other end at *B*. The electron that leaves the conductor, then, is not the same one that enters it. The effect is the same as if we stood up a row of dominoes and pushed over the first one in the line. It will push over the second, which then will tip over the third, and so on down the row. The movement down the line of dominoes far exceeds the rate of movement of any one domino. In a similar fashion, the effect of an electric current passes through a conductor from electron to electron.

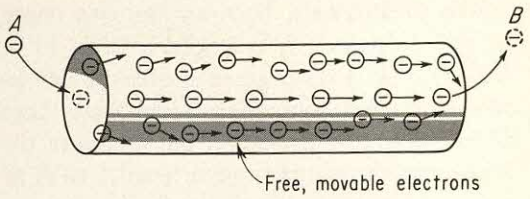


Fig. 9-1. The mechanism of electron migration in a metal rod.

9-2 Electrical Units

It will help our understanding of electricity to know some of the units by which we describe it. The *coulomb* (named after the man who, as we saw in the last chapter, discovered the law of electrostatic charge) is the unit used to measure the quantity of an electrical charge. When a coulomb of electrical charge flows by a certain point in a conductor, 6.2×10^{18} electrons have passed that point. When we have a coulomb of electric charge pass a point in a circuit during one second, we say one *ampere* of electric current is flowing. In other words:

$$\text{amperes} = \frac{\text{coulombs}}{\text{second}} \qquad (9.1)$$

Amperes are measured by an instrument called an *ammeter*. When the ammeter in an automobile reads "charge 5," it means that the generator is charging the battery at the rate of five amperes, or five coulombs a second.

A *volt* is a measure of the *energy* of electrons as they flow in a conductor. An electrical current occurs when electrons move from one point to another, and the energy that causes the electrons to move may be likened to potential energy, that is, electrons move from a position of higher potential energy to a position of lower potential energy. The energy change that causes an electric current to flow is spoken of as a difference of potential and is measured in volts. A *volt* is defined as that difference of potential which requires one joule of work to carry one coulomb of charge from one point to the other. Each time 6.2×10^{18} electrons, or one coulomb, move from a higher to a lower potential energy and in the process do one joule of work, a potential difference of one

volt is said to exist between the two energy levels. We can visualize this in Fig. 9-2, where 6.2×10^{18} electrons at A fall from A to B, a distance equivalent to one joule of energy and to a change in potential of one volt. According to the law of conservation of energy, the source of the work these electrons can do as they move from A to B is the difference in energy or potential between the electrons at A and at B. We can now put this relationship into an equation:

$$V = \frac{W}{q} \qquad (9.2)$$

where V is volts, W is joules, and q is coulombs. In a 12-volt storage battery the electrons have more potential energy when on the negative pole than they do on the positive pole and so they flow from the negative pole to the positive pole when we connect these two poles together with a wire. Each time one coulomb of electrons flows from the negative to the positive pole it loses 12 joules of energy, thus forming a 12-volt difference in potential.

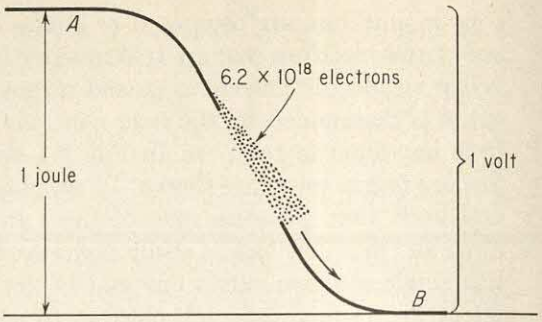


Fig. 9-2. A diagrammatic definition of a volt.

We know that electricity flows more easily through some substances than through others. For instance, copper is commonly used as a conductor primarily because it offers very little resistance to the flow of electrons. This resistance is measured in ohms; an *ohm* is the amount of resistance in a conductor that permits a drop in potential of one volt to produce a flow of one ampere of current (Fig. 9-3). Ohm's law, then, states:

$$\text{volts} = \text{amperes} \times \text{ohms} \qquad (9.3)$$

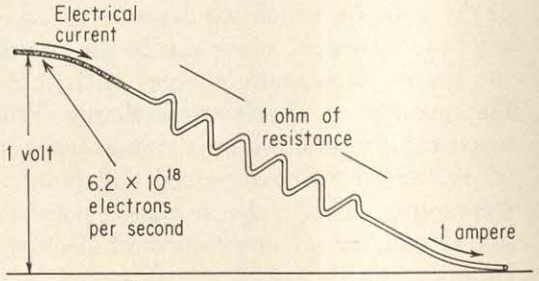


Fig. 9-3. A diagrammatic definition of an ohm.

Although many more electrical units are used by the scientist and the engineer, for our present needs the four defined above will be sufficient. If we relate these units to water flowing in a pipe, we can see more clearly what they are:

Electrical unit	*Water pipe analogy*
volt	water pressure
coulomb	gallons of water
ampere	gallons of water per second
ohm	resistance to the flow of water

From these comparisons, we may draw a further analogy from Ohm's law. If there is a large resistance to flow, then a high water pressure is needed to cause a given quantity of water to flow per second. Or, if the resistance is low, the same high pressure will cause a large amount of water to flow per second, just as a high voltage will produce a large amount of current (amperes) when the resistance of a wire, measured in ohms, is low.

9-3 Electrical Energy and Power

All of us use electricity in one way or another. As long as we pay our light bill, we usually give little attention to what

we actually buy from the power company. By analyzing our electric bill, let us see what we get for our money. We are usually charged with so many *kilowatt-hours*. The prefix "kilo," you will recall, means one thousand, and if we have used 400 kilowatt-hours, we have used 400,000 watt-hours, or 400,000 watts operating for one hour. What, however, is a watt?

On each of our appliances—toasters, radios, washing machines, hedge trimmers, and what have you—there is a cryptic message we seldom heed that gives the number of watts the item uses at a particular voltage. Many light bulbs are marked 50W 120V, which means they are designed to operate efficiently on 120 volts, and they would glow very dimly or not at all if they were connected to an automobile electrical system providing only 6 or 12 volts. Fortunately for most of us, 120-volt and 6-volt bulbs ordinarily have bases of different size so that it is impossible to interchange them. But, what about the 50W (50-watt) notation? A watt is a measure of the rate at which energy is used, and the number of watts a lamp uses depends upon both volts and amperes. We already know that:

$$\text{volts} = \frac{\text{joules}}{\text{coulomb}}$$

and:

$$\text{amperes} = \frac{\text{coulombs}}{\text{second}}$$

If we multiply volts by amperes, we get:

$$\text{volts} \times \text{amperes} = \frac{\text{joules}}{\text{coulomb}} \times \frac{\text{coulombs}}{\text{second}}$$

$$= \frac{\text{joules}}{\text{second}}$$

Joules per second (the rate of doing work) you should recognize as a unit of power. We give this power unit its own special name of *watt*. A watt, then, is one joule per second. So we can write:

$$\text{watts} = \text{volts} \times \text{amperes} = \frac{\text{joules}}{\text{second}} \qquad (9.4)$$

Our 50-watt bulb, therefore, uses 50 joules of energy each second it is lighted with 120 volts of potential difference. Since both the watt and horsepower are units of power you may be interested to know that 746 watts equals one horsepower. A 1500-watt electrical appliance then uses energy at the rate of about two horsepower.

Now back to the term *watt-hour*. This simply means that we use energy at a rate which averages a joule per second for one hour. It is an amount of work done or energy used. If we used the 50-watt lamp for an hour it would use 50 watt-hours of energy. Ten 50-watt lamps burning for two hours would use 1000 watt-hours, or 1 kilowatt-hour. When we use

a kilowatt-hour of electricity, we use a total of 3,600,000 joules

$$(1000 \frac{\text{joules}}{\text{second}} \times 60 \frac{\text{sec}}{\text{min}} \times 60 \frac{\text{min}}{\text{hr}}).$$

Thus, when we buy a kilowatt-hour of electricity, we purchase 3,600,000 joules of energy; the power company, if we are to be precise, bills us for the amount of energy, not the amount of power, we use. The means by which the electric company generates this energy will be discussed later in this chapter.

9-4 Magnetism

Nearly everyone at some time or other has played with a small magnet and found that it attracts iron or steel with an invisible, mysterious force but that it does not attract paper, glass, wood, and most other metals. In fact, the magnet's attractive force will pass through pieces of these materials and still affect iron and steel. A compass placed near a magnet will swing madly around and disregard its normal duty of pointing northward. In studying the force that causes these phenomena, we will get an insight into some of the more interesting topics in physical science—the structure of the atom, the nature of light, and the earth's magnetic field.

As we have already seen, lodestone, or magnetite (iron oxide), is a magnet and it can be detected with the aid of a compass. If a piece of lodestone is supported on a thread it will act as a compass and align itself in a north-south direction. Natural magnets, such as lodestone, are not very powerful, but man-made magnets of an alloy containing aluminum, nickel, and cobalt (alnico) are capable of lifting at least a hundred times their own weight. This type of magnet has become almost indispensable in telephones, radios, radar, and other electrical devices. A few other alloys can be magnetized, but most substances have limited magnetic properties. A recently developed alloy of manganese and bismuth makes a magnet ten times stronger than alnico; this new material seems to hold its magnetism better and can be fashioned in a form that can easily be cut and worked into intricate shapes. Such a magnet may open up entirely new possibilities in the fields of electrical power and communication.

A compass is a small magnet that may be constructed to swing either horizontally or vertically (Fig. 9-4). It aids us in traveling because it is attracted by the earth's magnetic field; the end of the needle that points northward is called a north (N) or north-seeking pole and the other end is called a south (S) or south-seeking pole. Unlike an object containing static electricity that may have either a positive or a negative charge, a magnetized object always contains both N and S poles. It is impossible to produce a single magnetic

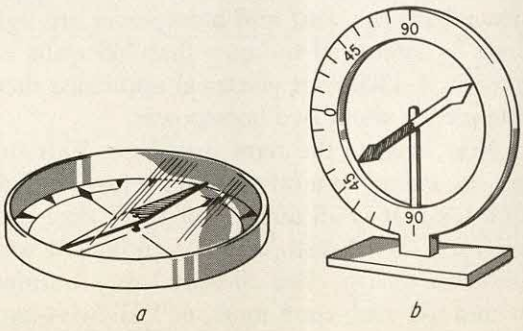


Fig. 9-4. Two different types of compasses.

pole. If we tried to isolate the N pole from the S pole by cutting a bar magnet in two, we would find that we have a new S and a new N pole formed at the cut (Fig. 9-5). No matter how many times we cut the magnet in two, we will always have a magnet with both N and S poles. On the other hand, magnetic poles are similar to electrostatic charges in that unlike poles attract each other and like poles repel each other. The force of repulsion can easily be shown by placing two bar magnets on top of each other, with their like poles together (Fig. 9-6). With suitable guides to hold the upper magnet in position, its weight may be fully supported by the magnetic forces between the two bars.

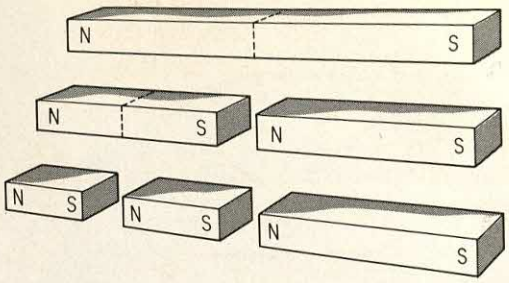


Fig. 9-5. Cutting a magnet in two produces two new poles.

Coulomb discovered that *the force of attraction between an N and an S magnetic pole is directly proportional to the product of their pole strength and inversely proportional to the square of the distance between them.* Can you write out the equation which might represent this relationship before you read further? This relationship appears very much like the one we encountered in Chapter 8 (Equation 8.1) concerning the force of attraction between electrostatic charges. In addition, it reminds us of Newton's law of universal gravitation (Equation 3.8). Suppose we compare these three kinds of forces—gravitational, electrostatic, and magnetic—and note their similarities and differences.

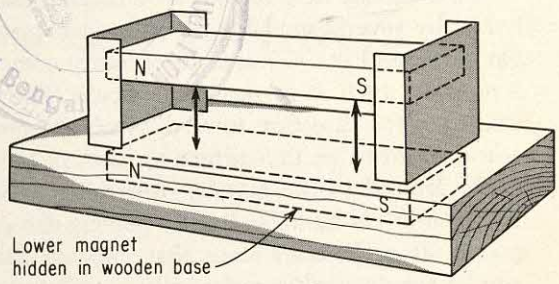


Fig. 9-6. One magnet supporting another by the repulsion of like poles.

In recent years scientists have attempted to find a single theory that would explain all three of these forces. Since they all have field effects—that is, they exert themselves through space—scientists hope to discover a theory that will explain why they have similar relationships, such as that of the inverse square. Although Einstein in his later years worked on a unified field theory, this problem still constitutes one of the most formidable questions facing the scientist today.

Gravity	*Electrostatics*	*Magnetism*
Attraction only	Like charges repel, unlike charges attract	Like poles repel, unlike poles attract
$F = G\frac{m_1 m_2}{r^2}$	$F = k\frac{q_1 q_2}{r^2}$	$F = \mu\frac{m_1 m_2}{r^2}$
Exhibited by all substances at all times	Exhibited by all charged substances when properly insulated	Exhibited strongly by iron, steel, cobalt, nickel, and certain alloys
No methods of shielding or stopping its effect	Conductors shield or destroy its effect	Its effect altered only by another magnetic substance
Gravity is inherent in any mass	Charges can be moved by a conductor	Magnetic poles are fixed in an object
Depends upon mass	Depends upon electrical charge	Depends upon lack of electrical symmetry in the atom, motion of electrons, and atomic orientation

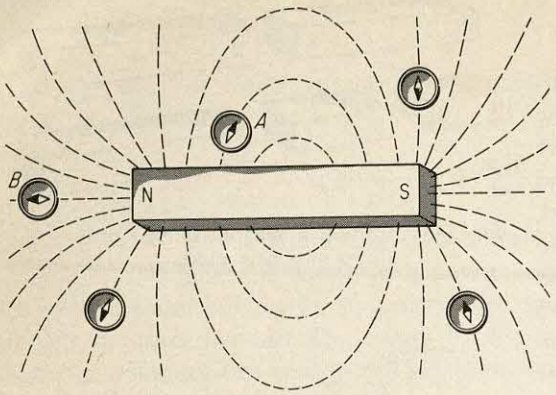


Fig. 9-7. Finding the magnetic field lines with a compass.

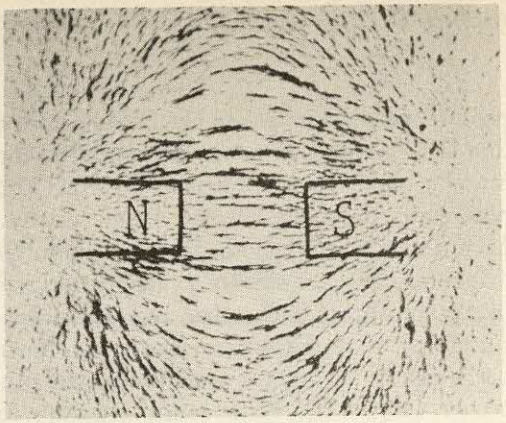


Fig. 9-8. Iron filings on a piece of paper lined up by the magnetic field between the north and south pole of a magnet.

The magnetic field formed by a magnet may be readily detected by several methods. In one, a bar magnet is laid on a flat surface (Fig. 9-7) and an ordinary compass is placed at a position, such as *A*, near the N pole. The compass is then moved in the direction toward which the needle points, and a line is drawn on the surface to mark its path. The line will curve around to the S pole, describing a symmetrical path between the N and S poles. By placing the compass at other points, as at *B*, more lines that indicate the magnetic field can be constructed to assist us in visualizing the effect of the magnet. A magnetic line of force is arbitrarily said to start at the N pole and end at the S pole. A second method of detecting a magnetic field is quicker and more dramatic. A magnet is placed beneath a sheet of paper covered with iron filings, which, after a gentle tap on the paper, will line up with the magnetic lines of force (Fig. 9-8).

What is the difference between a magnetized and an unmagnetized bar? In order to answer this question we must account for the following facts:

1. A magnet always has two poles.

2. A magnet, when broken in two, has a *new* set of magnetic poles.

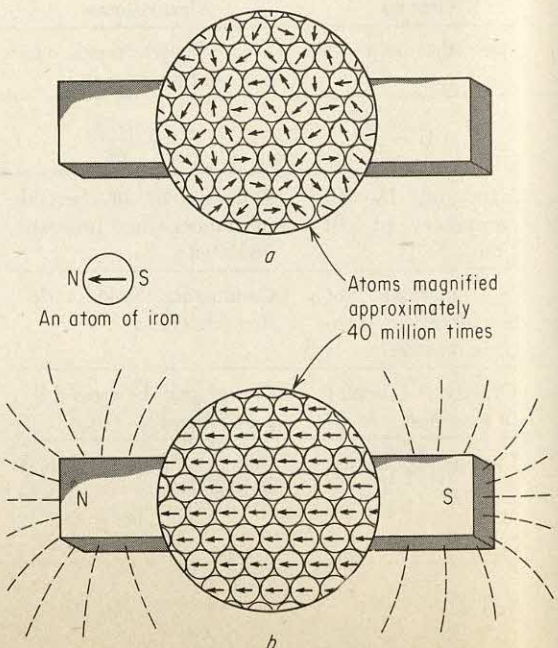
3. Heating or jarring a magnet weakens or destroys its magnetism.

4. Soft iron is easily magnetized by placing it in a magnetic field.

5. A piece of iron becomes slightly longer when it is magnetized.

A good way to explain these facts is to consider a magnet to be a collection of many tiny magnets, formed by groups of atoms or by the individual atoms themselves. (We shall see later how a single atom itself may act as a magnet.) In an ordinary piece of iron these little magnets are disordered, as shown in Fig. 9-9a. When the iron is magnetized, however, these tiny magnets line up with their N poles oriented in one direction and their S poles in the opposite direction (Fig. 9-9b). This theory of the structure of the magnet explains

Fig. 9-9. Diagram of the elementary magnets in a piece of iron.



why we always find two poles, and why cutting a magnet in two results in two new poles at the cut. It also explains why heat destroys magnetism—at high temperatures the tiny magnets become free to swing away from their oriented positions. When a piece of iron is placed in a magnetic field, the tiny magnets in it will arrange themselves in an orderly fashion and thus form a magnet. A piece of iron may be magnetized and de-magnetized over and over, with no apparent change in the composition of the iron. Finally, since the tiny particles in a magnet are apparently not exactly spherical, if the iron atoms are all oriented, as in Fig. 9-9b, they will take up slightly more space than if they were randomly oriented.

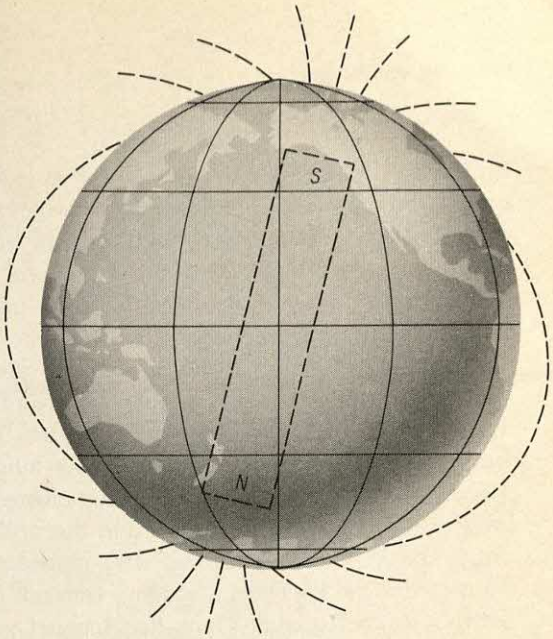


Fig. 9-10. The earth as a huge magnet.

9-5 The Earth's Magnetic Field

The earth is a huge magnet, but why it is one has baffled scientists for a long time, because at the high temperatures that apparently prevail inside the earth, its iron core should have difficulty serving as a permanent magnet. The most commonly accepted theory seems to be that deep internal currents flowing in the outer core of the earth give rise to the earth's magnetism, but the source of energy for these movements is still a major problem. Several other theories have been proposed which would make interesting reading. A full explanation of the earth's magnetism remains to be found.

The axis of the earth's "magnet" is displaced 12 degrees from the geographical axis (Fig. 9-10) and acts very much as though it were a large bar magnet buried within the earth. The north pole of the earth's magnet (which is actually an S pole, since it attracts the N pole of a compass) is located beneath a point about a 1000 miles north of Hudson Bay, 850 miles short of the North Pole. Consequently, the needle of our horizontal compass will not point exactly toward the North Pole except in a narrow north-south region extending through Lake Superior and the southeastern states (Fig. 9-11). When we measure the direction of its magnetic field

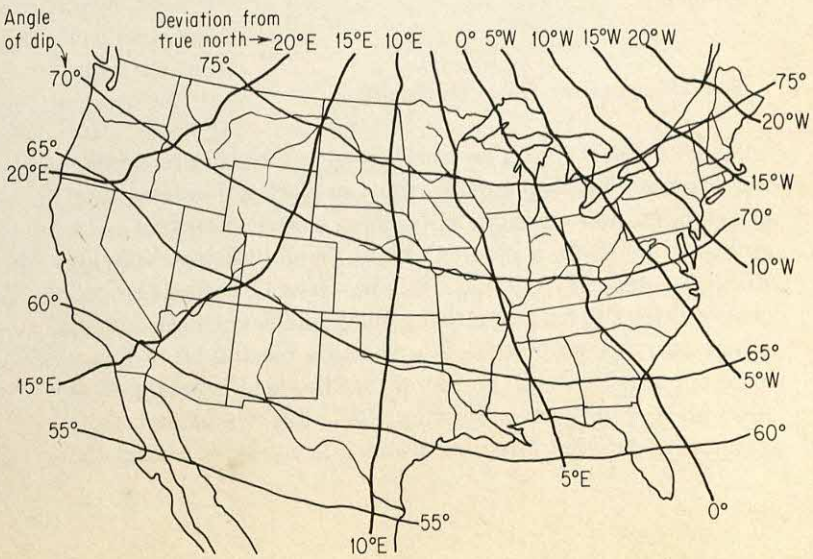


Fig. 9-11. A magnetic map of the United States showing the deviation of the compass from true north and the angle of dip.

lines with dip-needle compass (Fig. 9-4b), the lines are horizontal only near the earth's equator. In some parts of the United States, these lines point downward at angles that range between 60 and 75 degrees from the horizontal.

The earth's magnetism seems to be an important factor in producing the aurora borealis. Where the earth's magnetic lines point downward 80 to 85 degrees from the horizontal —at places near the poles, in other words—the aurora is visible on over 80 per cent of the nights. The earth's magnetic poles apparently move slowly over long periods of time. Recent studies of the residual magnetism in ancient rocks indicate that 700 million years ago the north magnetic pole might have been in Arizona. Any magnetizable rock, such as an iron ore, becomes a magnet through the effects of the earth's magnetic field. Once this magnet is formed, assuming that its position is not changed, it will remain in the direction of the original magnetic field that induced the magnetism in the rock. A small amount of residual magnetism will, in a sense, be frozen in the rock. The north magnetic pole seems to have moved slowly across the Pacific to Japan, then northward to Siberia, and finally to its present position.

The direction of the earth's magnetic field seems to have been reversed at least twice in the course of geologic time. Magnetic north lay deep in the southern hemisphere between 980,000 and 1,900,000 years ago and prior to 3,400,000 years ago. Since the geographic poles and the magnetic poles supposedly remain near each other, this hypothesis might explain the radical changes in climate that have occurred over the face of the earth during past geological ages. Arizona might at one time have been in a polar, frigid region, and later Japan and Siberia may have moved into the polar region. At any rate, this mystery is another one of the many interesting, unsolved problems in physical science.

While we may take the earth's magnetic field for granted, present evidence indicates that Venus has little or no magnetic field surrounding it. The effect of a planet's magnetic field on charged particles coming into the planet makes the study of planetary magnetism an important field of space research.

9-6 Magnetism from Electricity

Most of the magnets in commercial use today are temporary magnets that are energized by an electrical current; one of the important characteristics of an electrical current is its ability to generate a magnetic field. While talking with his students after class one day in 1819, Hans Christian Oersted observed for the first time that a magnetic field exists around a wire carrying an electrical current. By placing an ordinary compass near the wire, he found that the field is arranged in concentric circles around the wire (Fig. 9-12), with its strength depending upon its distance from the wire and the

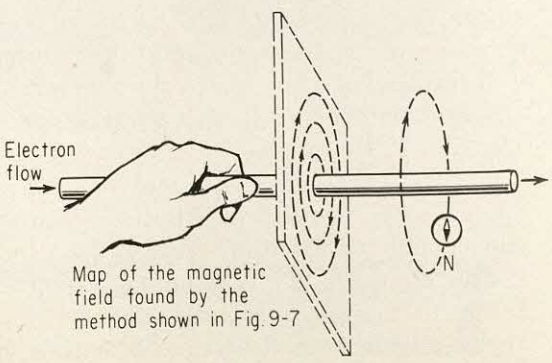


Fig. 9-12. The magnetic field around a wire in which there is a flow of electrons.

amount of current flowing through the wire. The direction of the field can be determined by the so-called left-hand rule: When you grasp the wire with your left hand, with your thumb pointing in the direction of the flow of electrons, your fingers point in the same direction as the N pole of a compass when it is placed in the magnetic field.

If we wind a wire into a coil and pass a current through it (Fig. 9-13), it acts like a magnet. In fact, since the fields around each wire reinforce each other, an intense magnetic field may be formed. If a piece of iron is placed inside the coil, a still stronger magnet, called an *electromagnet,* is produced. In this case, the magnetic field of the electric current produces a temporary magnetism in the iron core which complements and strengthens the whole magnetic effect. Since soft iron is used, the induced magnetism is only temporary and it acts like a magnet only while the current is flowing through the coil. Here is an amazing, yet extremely useful, peculiarity of the electron. Scientists, as we have seen, believe that the electron is a tiny mass containing a single negative charge and that it accounts for the negative charge of such substances as a fur-stroked rubbed rod. But, since this charged rod shows no magnetism, we must conclude that it is the *movement* of the electron that generates a magnetic field. Thus, a moving electron, whether it is carried by a conductor or not, creates both electrostatic and electromagnetic effects, which, as we shall soon see, means that its energy manifests itself in a rather complex manner.

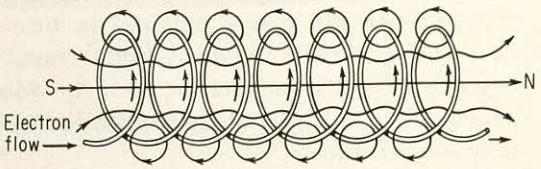


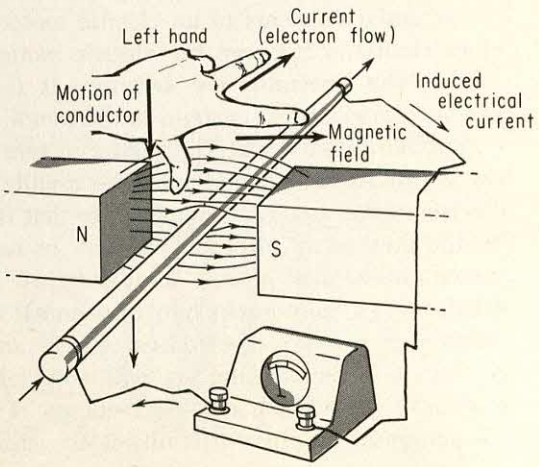
Fig. 9-13. A magnetic field is produced by an electrical current flowing through a coil of wire.

9-7 Electricity from Magnetism

We saw above that a moving stream of electrons produces a magnetic field around a conductor. What would happen if a conductor, such as a copper wire, were moved through a magnetic field? A simple experiment will give us the answer. Let us place two unlike magnetic poles a short distance apart so that an intense magnetic field is concentrated between them. Then let us move a copper wire, with an ammeter attached to detect a flow of current, quickly down between the two poles, in effect "cutting through" the magnetic field (Fig. 9-14). When we exert enough force to move the wire downward through the magnetic field, the ammeter will indicate that a current is flowing in the wire. Note that the direction of the magnetic lines, the movement of the wire, and the flow of electrons in the wire are all mutually perpendicular to one another. If we push the wire upward, the electrons in the conductor will move in the opposite direction. When we move the wire up and down at various speeds, we will note that the faster the movement, the greater the amount of current produced in the conductor.

This process is the basic principle of the generator or dynamo. In an actual generator, however, the moving wire is replaced by a coil in the form of a loop, so that the maximum

Fig. 9-14. Producing a flow of electrons by a magnetic field.



efficiency of the mechanism can be obtained (Fig. 9-15). Instead of bar magnets, powerful electromagnets produced by electrical current flowing through a coil, as shown in Fig. 9-13, are used to create the magnetic field. The current produced by a conductor moving through a magnetic field is called an induced current, and it is by the method described above that the major portion of our electrical energy is generated. Since the power required to turn the coil of the generator is usually obtained from steam, waterpower, internal combustion engines, or even the wind, we find here an important example of the conversion of one form of energy into another.

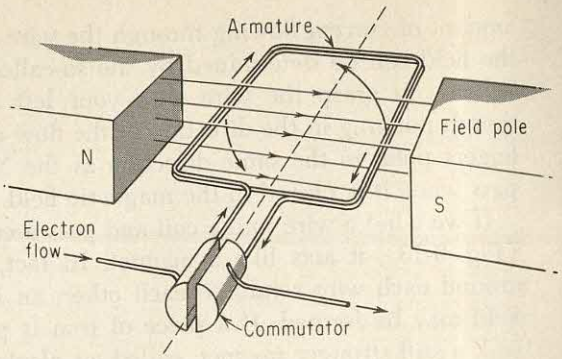


Fig. 9-15. The principle of the electrical generator.

9-8 Mechanical Energy from Electrical Energy

In the last section we saw how mechanical energy is converted into electrical energy. Now let us see how electrical energy is transformed into mechanical energy. The first step is to examine the effect of a magnetic field upon a conductor carrying an electrical current. You will recall that such a conductor has its own magnetic field around it. Figure 9-16 illustrates the interaction of the magnetic field around the conductor and an external magnetic field. As an aid in understanding this experiment, think of the magnetic field as consisting of taut rubber bands. Below the wire in the illustration the two fields act together and are strengthened, while above they are opposite in direction and produce a weakened magnetic field. As a consequence, the field exerts a net repulsive force on the wire, pushing it upward through the weakened field. Hence, in Fig. 9-16 the conductor will be forced upward by the interaction of the two magnetic fields.

Let us now replace the single conductor in Fig. 9-16 with a loop of wire that can be rotated, such as the one in Fig. 9-15. If we attach the wire from the commutator to a battery and produce a current through the wire loop, the resultant forces on the loop will cause it to rotate, and we thus have the essential elements of an electric motor. There is an obvious similarity between the electric motor and the generator: In the generator the armature is turned in order to produce an electrical current; in the motor, electrical current is supplied to the armature in order to turn it. As in the generator, the fixed magnetic poles are usually replaced with an electromagnet, thus creating a device that is actuated entirely by the flow of an electrical current. In sum, we have converted mechanical energy into electrical energy and electrical energy back again into mechanical energy. Since the production and use of electrical energy are so important in our modern way of life, we will now take up briefly the mechanism by which electrical energy is transported from the generator to where it is ultimately used.

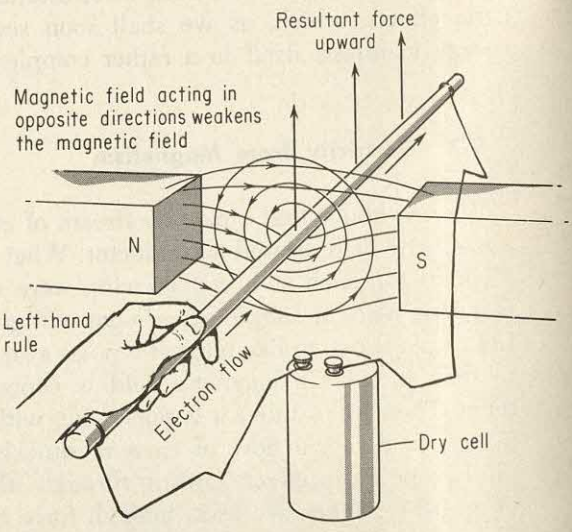


Fig. 9-16. The interaction of two magnetic fields produces mechanical energy from electrical energy.

9-9 Direct and Alternating Currents

Up to this point, we have been concerned primarily with the flow of electrons in a conductor without bothering about the continuity of the current. Since many useful devices, however, are made possible by altering the flow of electrons in an electrical current, we will want to know something about this modification of the current. The current produced by batteries and by certain types of generators is *direct current* (DC), which means the electrons in the conductor flow continuously in one direction. This steady flow of electrons may be measured by voltage and amperage, and it establishes a uniform, fixed magnetic field around the conductor so long as the current is flowing. The electrical current in our automobiles, flashlights, and battery-operated radios is direct current.

Most of our electricity, however, is produced by generators that move the electrons first one way in the conductor and then the other way. In the ordinary flow of electricity supplied to domestic users in the United States, this change in direction is made 120 times per second, in what is called a 60-cycle *alternating current* (AC). The reversal occurs so quickly that we ordinarily do not see its effect upon our light bulbs. Many electrical motors are constructed to operate on the alternating flow of electrons in alternating current. A clock motor, for instance, runs at a constant speed only so long as the number of cycles per second in the current supplied to it is constant. The electric power company can speed up or slow down our clocks by simply increasing or decreasing the number of cycles per second in the current that it produces, but fortunately, the number of cycles in AC is very rigidly controlled.

There is a big advantage in transporting electrical energy in the form of an alternating current instead of a direct current, because the voltage of alternating current can be simply and efficiently reduced or increased. Direct current, on the other hand, can be reduced in voltage only by passing it through a resistance, and this is costly because part of the electrical energy is dissipated as heat; furthermore, the voltage of direct current cannot be stepped up in a simple manner. You might think that increasing the voltage of alternating current will increase the energy and thus refute the law of conservation of energy, but it does not, for when AC voltage is increased its amperage is proportionately decreased. Since electrical power (watts) is the product of volts and amperes, it is obvious that the voltage can be increased if the amperage is decreased, so that the product of the two remains constant and the law of conservation of energy maintained.

Let us now examine the method by which the voltage of alternating current can be changed—a method that involves a simple electromagnetic process. If we place a coil of wire *A* (Fig. 9-17) alongside a second coil *B* and then pass an

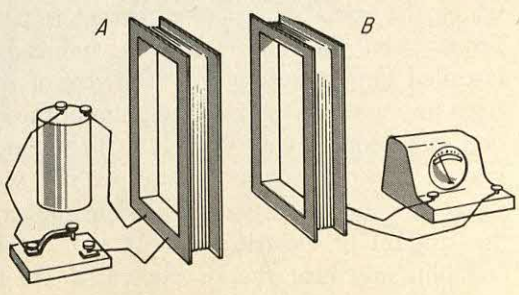


Fig. 9-17. Producing an electrical current by induction.

electrical current through coil *A*, we will see the ammeter needle move, showing that a current is induced in coil *B*. The current in *B*, however, flows for only an instant, because the needle of the ammeter soon falls back to zero. Now let us turn off the current flowing through coil *A*. Again, the ammeter shows that a current is produced in coil *B*, but it now flows in the opposite direction. Each time we start or stop the current in coil *A* an induced current is produced in coil *B*, the current flowing first one way, then the other way. What is the reason for this effect? When an electrical current starts to flow in a conductor, the magnetic field that forms around it grows outward from the conductor in a widening circle. As the current is started through coil *A*, the magnetic field moves outward and part of it is cut by coil *B*, producing an induced current in it. When the current in coil *A* is stopped, the magnetic field around this coil contracts or collapses onto the conductor. Some of this shrinking magnetic field is again cut by coil *B* and induces a current in it in the opposite direction. In this experiment we are witnessing the conversion of electricity to magnetism and back to electricity each time a current starts or stops in coil *A*. The only moving part in our apparatus is a switch for turning on and off the current to coil *A*. You will recognize this as the same process discussed in Section 9-7.

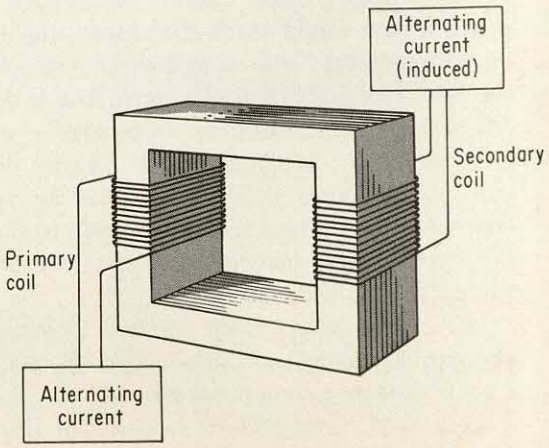
Instead of starting and stopping the current in coil *A* by a mechanical switch, we might supply an alternating current to this coil. Now the current starts, stops, reverses its direction, stops, reverses again, and so on. Each time there is a change in the current, there will be a current induced in coil *B*. Here, then, is the essence of a *transformer*. If an equal number of turns of wire are in coils *A* and *B* in Fig. 9-17, the voltage induced in coil *B* will be equal to the original voltage in coil *A*. However, by varying the number of turns of wire in the two coils, we can change a particular voltage into almost any voltage we wish. The rule by which we may determine the voltages in the two coils is:

$$\frac{\text{number of turns in } A}{\text{number of turns in } B} = \frac{\text{voltage in } A}{\text{voltage in } B}$$

The ability of AC to be changed to other voltages through transformers is an obvious advantage of this type of current over DC.

A transformer's efficiency is increased by placing both coils around a common piece of iron (Fig. 9-18). The coil through which the initial alternating current is passed is called the *primary coil;* the coil in which the induced current is formed is called the *secondary coil.* The size of transformers varies from tiny ones that fit into the palm of the hand to huge ones that may weigh many tons. It is interesting to note that the law of conservation of energy is not violated here. The amount of electrical energy used in the primary coil equals the amount of electrical energy put out by the secondary coil plus any heat that is generated. As the voltage is in-

Fig. 9-18. A simple transformer.



creased, the amperage is decreased, so that the product of volts times amperes, or watts, is always constant. Even though the process of converting one voltage to another is highly efficient, a small fraction of the electrical energy is always converted into heat energy. If we account for both the electrical energy and heat energy involved, no energy is lost or created. In other words, the work put into the system equals the work put out by it. In the form of an equation this becomes:

$$\text{Power in primary} = \text{Power in secondary}$$

$$P_p = \text{volts}_p \times \text{amperes}_p = P_s = \text{volts}_s \times \text{amperes}_s$$

A transformer is a crucial link in transporting electrical energy. Generating power plants are often located at a considerable distance from the cities and industries that use the energy they produce. Since less electrical energy is lost if it is conducted across country at high voltage than at low voltage, you will always see transformers at a power plant for the purpose of stepping up the voltage of the electricity before it is sent out through the transmission lines. On the consumer's end, the voltage is then stepped down, usually to 110 volts, through several transformers.

9-10 Series and Parallel Circuits

A brief introduction to electrical circuits will give us an opportunity to summarize some of the concepts of electricity we have encountered in this chapter. It will also enable us to examine electrical circuits in practical situations that most of us frequently encounter.

In a *series circuit*, all the electrical current flows through the whole system. A good example of a series circuit is that of an ordinary storage battery (Fig. 9-19), which consists

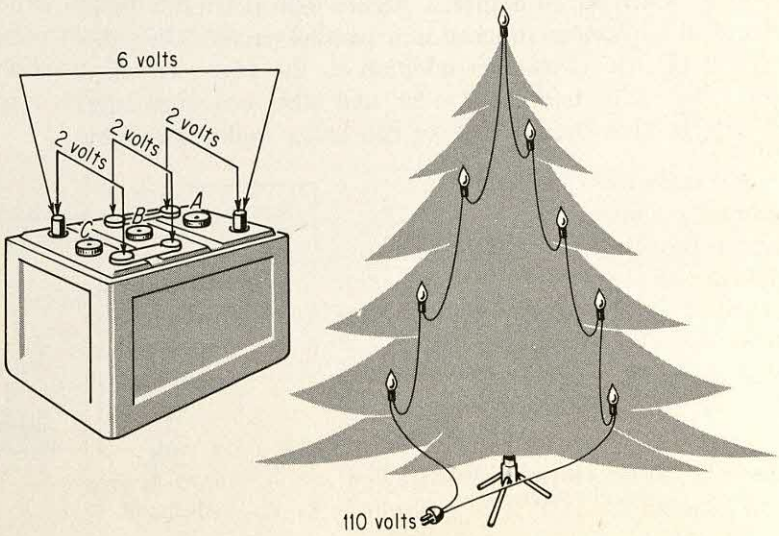


Fig. 9-19. Examples of series circuits: a storage battery and ornamental Christmas tree lights.

of three separate cells, each capable of producing 2 volts of electricity. The process by which each cell generates this voltage is primarily of a chemical nature, which need not concern us at this point. The negative pole supplies electrons that flow through the outer circuit and reenter the cell at the positive pole. From cell *A*, electrons put out at 2 volts enter cell *B* by way of *B*'s positive pole, emerge from *B*'s negative pole with 2 more volts of potential energy, and finally emerge from cell *C* with a total of 6 more volts of energy than they had at the positive pole of cell *A*. The voltage of a battery, therefore, is the sum of the voltages of the cells that are connected in series. An automobile's 12-volt battery has six such cells in series.

What happens to the current, or amperage, in a series circuit? Since an electrical current is a flow of electrons and *all* the electrons must eventually flow through each cell in a battery, we find that exactly the same number of amperes of current flows through each cell. Suppose we use some electrical energy in a series circuit, say in a string of Christmas tree lights (Fig. 9-19). All the electrons in the current go through each light, hence the amperage through each light will be identical. If 110 volts are supplied to the circuit, a portion will be used by each light, and as the electrons flow through the light they will suffer a drop in potential or voltage. Thus, in Fig. 9-19 there is a drop of 110/8 volts for each light. For the string of lights to operate efficiently, the voltage drop across each bulb should be identical to the voltage for which it is designed. If one of these lights should "burn out," no current can pass through it, and all the lights will go out.

A more common type of circuit is the *parallel circuit*. In it each light or appliance uses up electrical energy independently of the others. Our houses are wired in parallel circuits, and in them all lights and appliances operate on 110 volts, with each of them using as much amperage as it may require: A 100-watt light uses almost one ampere, a motor may require 5 amperes, and an electric stove may need as much as 50 amperes. Figure 9-20 illustrates three electrical appliances attached in a parallel circuit. There are several electric circuits in addition to the two mentioned here. Our radio, television, radar, and other communication devices involve circuits that we can better understand after

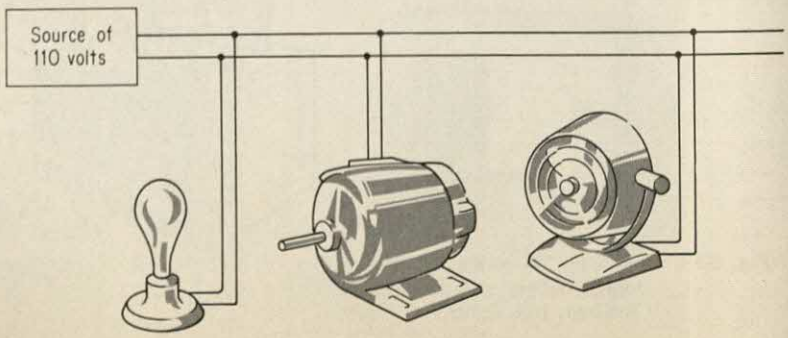


Fig. 9-20. Home appliances attached in a parallel circuit.

we become acquainted with the properties of wave motion that will be discussed in the next chapter.

9-11 Summary

Current electricity is a flow of electrons in a conductor. The quantity of an electrical current is measured in coulombs, one coulomb being 6.2×10^{18} electrons. One coulomb of charge moving past a given point in a conductor in one second is a current of one ampere. The electrical potential is measured in volts, one volt being one joule per coulomb.

If one ampere flows through a conductor at a potential difference of one volt, the conductor has one ohm of resistance. Ohm's law states that the voltage drop of an electric current is amperes times ohms.

Electric power is measured in watts. A watt, the product of volts and amperes, is one joule per second.

Magnetism is similar to electrostatics in that it consists of two opposite poles, the north and south poles. The force of attraction and repulsion between opposite and between like poles obeys a law identical in form to Coulomb's law for electrostatic charges. Magnetism is exhibited mostly by certain metals such as iron and steel.

The earth's magnetic field has a south pole near the geographic North Pole, and a north pole near the geographic South Pole.

A movement of electrons always generates a magnetic field at right angles to their direction of movement. If a magnetic field is cut by a conductor, an electric voltage is induced in the wire. It is upon these two principles—the conversion of electrical energy into a magnetic field and a magnetic field into electricity—that most of our modern uses of electricity depend. Generators, motors, and transformers are all based on these two principles.

Direct current is the passage of electric current in one direction only, while alternating current involves the periodic reversal of the flow of current. Only alternating current can be stepped up or down in voltage by means of transformers.

Most electricity is used in parallel circuits, although occasionally there is an advantage to using a series circuit.

EXERCISES

A. KEY TERMS TO REMEMBER

alternating current
ampere
compass
coulomb
Coulomb's law of magnetism
dip-needle compass
direct current
electrical current
electromagnet
generator
induced current
magnetism
N pole
ohm
parallel circuit
S pole
series circuit
transformer
volt
watt

B. QUESTIONS ABOUT ELECTRICITY AND MAGNETISM

1. List three important effects produced by an electrical current that are useful in modern civilization.

2. Ten identical lights are in series in a string of Christmas lights attached to a 110-volt source of electricity. What voltage is across each light? What would be the voltage if the lights were in a parallel circuit? In each case what will happen if one of the lights burns out?

3. Explain why the earth's magnetic pole near the north geographic pole is actually a south pole.

4. Make a list of the similarities and differences between electric charges and magnetic poles.

5. When do electric charges and magnetic poles have no effect upon each other?

6. What are three common forces that exert their effects through space?

7. When an electrical current travels through a conductor, would you expect to find the free, moving electrons distributed uniformly throughout the conductor, or concentrated in some parts of the conductor?

8. List some of the additional information that scientists would like to have in order to explain the origin of the earth's magnetism.

9. Where on the face of the earth will the north pole of a dip-needle compass point (a) straight downward, (b) straight upward, (c) horizontally?

10. What will a horizontal compass needle do when placed directly over the north magnetic pole?

11. Describe briefly the relationship between electricity and magnetism that permits us to convert mechanical energy into electrical energy.

12. Which compass in Fig. 9-7 indicates the field incorrectly? What might cause such an indication?

C. PROBLEMS ABOUT ELECTRICITY AND MAGNETISM

1. How many electrons move past a point in a conductor per second when there is a current of 5 amperes? Ans. 31.0×10^{18} electrons.

2. If a wire has a resistance of 50 ohms, how many amperes of current will be produced if there are 100 volts potential difference between the two ends of the wire?

3. What is the maximum wattage that can be used on a 110-volt circuit that is protected with a 20-ampere fuse?

4. How many amperes flow in a circuit when a 1500-watt heater is plugged into a 110-volt line? Ans. 13.6 amperes.

5. If 2200 volts AC are attached to a transformer's primary coil that has 2000 turns, what voltage is induced in a secondary coil that has 500 turns?

6. If we had two 100-watt 110-volt lamps, but only had available a 220-volt line, how could we connect these so the lamps would work effectively?

7. If we have a ½-horsepower 110-volt vacuum cleaner, what current will it take?

8. If electricity costs 5 cents per kilowatt-hour, how much will the electricity cost to run the vacuum cleaner described in problem 7 for one hour?

9. What maximum current would flow in a 110-volt house circuit to which was connected a 1500-watt toaster, a 600-watt coffeemaker, and a 1200-watt frypan?

10. In problem 9, what total horsepower would be involved?

Ten

Wave Motion

Although we are usually not aware of it, most of the information we receive from sources outside ourselves comes to us through wave motions. Of our five senses, the two most important—hearing and sight—depend on our perception of sound waves and light waves respectively. The aesthetic strains of a symphony orchestra, the murmur and clatter of a busy city, and the sharp clap of a thunder storm all result from a movement of energy in the form of sound waves from the source to our ears.

The colors of the rainbow, infrared, visible, and ultraviolet light poured out by the sun, radio waves, X rays, and gamma rays are all examples of electromagnetic waves. We have all experienced other types of waves—water waves, waves in a stretched violin string or piano wire, and perhaps even earthquake shocks. Some of these waves need no medium and can travel through a vacuum whereas the other waves mentioned here require the presence of matter, such as air, metals, wood, and so on, in order to be transmitted from one place to another.

In addition to the various qualities of wave motion, the physical scientist is deeply interested in the transmission of energy by waves because the behavior of waves serves as a key to the structure of matter itself. The inner structure of the atom, the power locked in the atom's nucleus, as well as the infinities of the universe have been revealed through man's increased knowledge of wave motion. It has been said that the studies made possible by only one of these kinds of waves—X rays—has revealed more about the structure of the natural world than any other tool the scientist has used. But before we go too far into the subject, let us examine the basic aspects of wave motion.

10-1 Wave Motion

A pebble tossed into a quiet pool of water will cause a series of waves on the surface of the pool. These waves start from the point where the pebble strikes the water, move outward in concentric circles, and become weaker as the circles become larger. When they reach the edge of the pool, they are reflected back, again in concentric circles. From these water waves we can learn much about wave motion in general. In the first place, a water wave is produced by an original disturbance, namely the pebble depressing and breaking the surface. When a portion of the water is lowered below its rest position, gravitational forces (and others) tend to return it to its original level. As it returns, inertia carries it upward beyond its rest position and a crest or high spot is produced. These high and low points, as they move outward, carry with them the ability to rock a boat and we can see that the energy of the falling pebble has been conveyed to the boat through the waves. Wave motion is then a means of transmitting energy from one place to another.

If you and a friend were each to hold the end of a rope, as in Fig. 10-1, you could perform a simple experiment in wave motion. By snapping the rope up and down, you would send a wave out through the rope, and your friend would experience a tug or jerk in the rope as the wave reaches him. If your friend holds the rope tightly, part of the wave may be reflected and return to your hand, where you would experience part of the original energy that you put into the rope. This simple experiment reveals some of the basic properties of wave motion that we shall discuss in the next section.

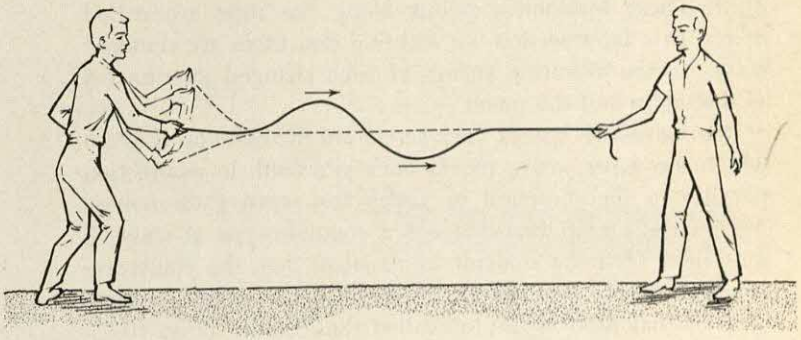


Fig. 10-1. The generation of wave motion in a rope.

10-2 Transverse, Longitudinal, and Complex Waves

The wave that travels along a rope when it is snapped is called a *transverse wave*, because each section of the rope moves crosswise or at right angles to the direction the wave moves. This action can easily be shown by tying a handkerchief to the rope and noting how it moves up and down,

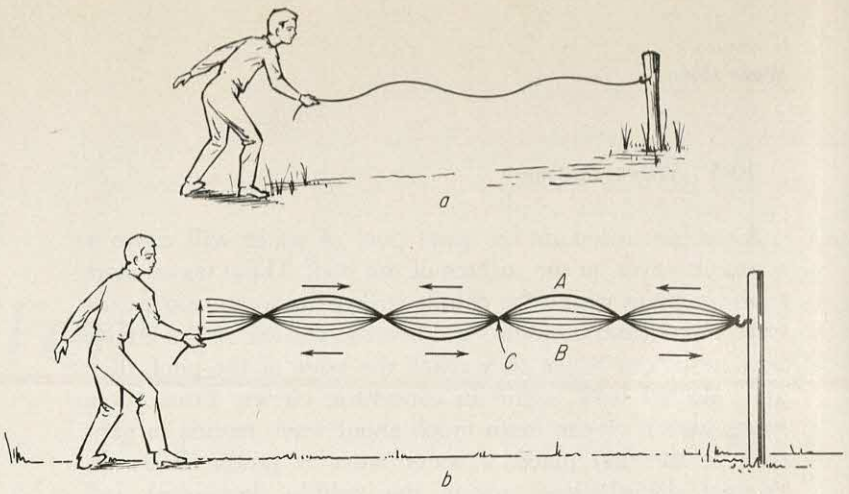


Fig. 10-2. The reflection of a train of waves resulting in standing waves.

rather than toward either end of the rope, as the wave passes. Even though the wave motion travels forward along the rope, the rope itself moves only up and down. If one end of a rope is attached to a pole and the other end is jerked up and down continuously, a series of waves will travel along the rope (Fig. 10-2a). As these waves are reflected, they may form *standing waves* (Fig. 10-2b), in which the points on the rope at the top of each wave loop, such as *A*, move up and down from *A* to *B* but the points such as *C* between the loops remain at rest, even though wave energy is being continuously passed through them both to the left and to the right. These motionless points along the rope are called *nodes*. In a later section we will find that there are standing waves in the vibrating strings of such stringed instruments as the violin and the piano.

In *longitudinal waves*, each section of the material through which the wave passes moves back and forth in a direction parallel to the direction in which the wave itself moves. Although a longitudinal wave is a common type of wave, it may be a bit more difficult to visualize than the transverse wave. A simple example of it can be observed in a spring, such as that used in the toy called the "Slinky" (Fig. 10-3). When this spring is stretched between the hands, it provides a suitable medium for longitudinal waves. If one hand is moved quickly toward, then away, from the other hand, the spring is compressed near the end, and this compression effect moves rapidly along the spring to the other hand, from which it is reflected back. Note that any given point on the spring moves backward and forward in a horizontal motion. Standing longitudinal waves can be formed in a manner similar to that which formed standing transverse waves (Fig. 10-2b).

One of the most complex types of wave is our old friend the water wave. As we stand on the bank of a lake we see the waves roll toward us as if the water itself were flowing

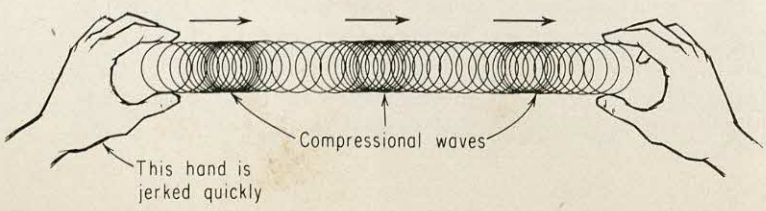


Fig. 10-3. Longitudinal waves produced in a spring.

toward the shore. Actually, however, the water remains almost stationary, and only the wave motion is moving. This can be easily verified if we watch the behavior of a cork floating on the surface of a lake when the waves are fairly high (Fig. 10-4a). We see that the cork not only moves up and down, but also forward and backward and upon close examination, that it follows almost exactly a circular path. As the wave passes through each point on the surface of the water, the water itself travels in a circular motion (Fig. 10-4b).

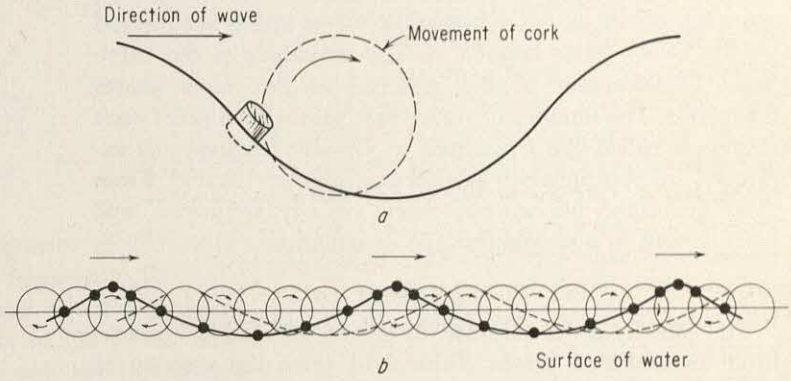


Fig. 10-4. A cork floating on water and a point on the water surface trace a circular motion as water waves pass by them.

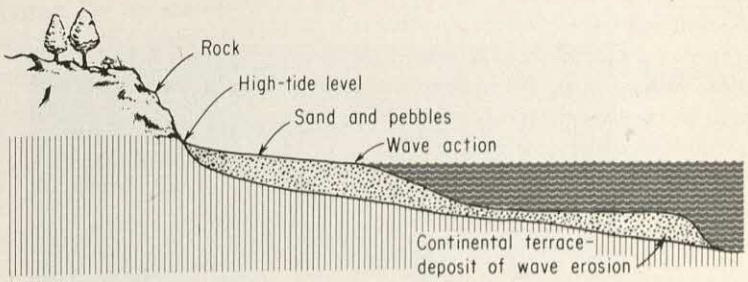


Fig. 10-5. A wave-cut cliff along an ocean shore line.

As an ocean wave approaches the shore, it tumbles forward because in shallow water the lower part of the wave is slowed down by contact with the ocean bottom; the upper part of the wave then breaks because it is ahead of the lower part. These waves, which are caused chiefly by the wind, assault the shore line with great force, and their eroding power is greatly increased by the sand and pebbles they carry along with them. Some of our most spectacular scenery is found along coastlines. A typical shore profile (Fig. 10-5) shows how the waves create a terrace that may extend outward from the shore for many miles. The terrace is composed of the rocks, pebbles, and sand that have been eroded away from the shore by the force of the waves.

10-3 Wave Length and Velocity

In the study of wave motion, there are several terms we need to know. A wave sent through a homogeneous material

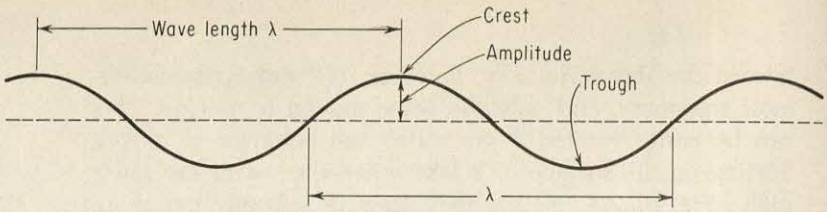


Fig. 10-6. The structure of a wave, showing wave length, amplitude, crest, and trough.

will travel at a uniform, constant velocity, and if a steady source of oscillation sends a continuous train of transverse waves through the material, the waves will appear as in Fig. 10-6. The *wave length* is the distance between corresponding points on two consecutive waves and is symbolized by the Greek letter lambda, λ. The *amplitude* is the maximum displacement of the material as the wave passes through it. The number of waves that pass a fixed point each second is called the *frequency, n.* Usually frequency is expressed in vibrations per second or cycles per second. From these definitions we can now relate velocity, frequency, and wave length in a simple formula or equation:

$$v = \lambda n \qquad (10.1)$$

The velocity of wave motion is greatly affected by the material in which it moves. Table 10-1 gives the velocity of sound waves, which are longitudinal waves, in various materials.

TABLE 10-1 VELOCITY OF SOUND IN VARIOUS MATERIALS

Air	1,130 ft/sec
Water	4,700 ft/sec
Wood (fir)	15,000 ft/sec
Iron	16,800 ft/sec
Beryllium	40,000 ft/sec

We have all noticed at some time that the whistle of a railroad train seems to drop in pitch when the train rushes past us and that the horn of an approaching car will also suddenly drop in pitch as the car passes. The change in pitch—called the *Doppler effect*—is caused by the movement of the source of the sound either toward or away from the observer. When the source moves toward the observer, the sound waves are crowded together into shorter wave lengths than normally, and since the frequency of sound waves determines the pitch of the sound, a decrease or increase in wave length decreases or increases the pitch. Suppose an automobile, traveling at a uniform velocity, sounds its horn (Fig. 10-7). The horn moves forward continually as it sends out waves, and to an observer standing at *A*, the

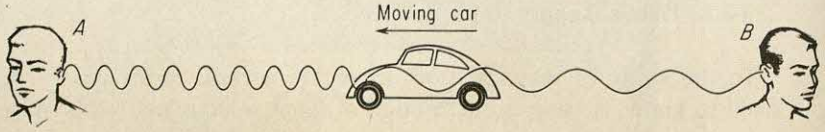


Fig. 10-7. The Doppler effect produced by an automobile horn.

sound waves coming from the horn will be pushed together and thus will have a higher pitch than they will have for the driver of the car. On the other hand, an observer at *B* will hear a lower pitch because the sound waves will be stretched out.

Let us apply Equation 10.1 to a typical situation where the Doppler effect is in operation. Since we know that the velocity of sound in air is 1130 ft/sec, we can calculate the frequency of a wave with a wave length of, say, 4.7 feet:

$$v = \lambda n$$

$$1130 \frac{\text{ft}}{\text{sec}} = 4.7 \text{ ft} \times n$$

$$n = \frac{1130 \text{ ft/sec}}{4.7 \text{ ft/cycle}}$$

$$= 240 \frac{\text{cycles}}{\text{sec}}$$

This sound wave, therefore, has a frequency of 240 vibrations per second. Now let us see what happens to the wave length and the frequency of this sound when the source moves toward the observer at a velocity of 125 ft/sec. Note that the velocity of sound through air will not be affected by the movement of the source, nor will the movement of the source affect the frequency at the source. But to the observer, the wave length will be shortened by the distance that the source moves during the formation of each wave. This wave takes 1/240 second to form (since 240 waves are formed per second) and the car during this time will move 125 ft/sec $\times$ 1/240 sec = 0.52 ft. (Do you recognize the use of Equation 3.1?) The new wave length will then be 4.7 ft $-$ 0.52 ft = 4.18 ft. The apparent frequency then, from Equation 10.1, is

$$n = \frac{v}{\lambda} = \frac{1130 \text{ ft/sec}}{4.18 \text{ ft/cycle}} = 270 \frac{\text{cycles}}{\text{sec}}$$

Hence, the frequency is increased from 240/sec to 270/sec, a difference that represents an increase in pitch of a whole note on the musical scale.

This particular example of the Doppler effect is only one application of this important principle. Astronomers use it in studying the light waves from stars to tell whether the stars are approaching or receding from us and at what rate. They can also use it to determine the rate of rotation of some of the nearer stars. Another application of the Doppler principle is the calculation of the speed of rockets. A fixed frequency radio transmitter carried aloft by a rocket appears to a ground observer to change frequency, thereby giving ground observers a continuous measure of its speed. Submarine detection and tracking devices also make use of this important effect.

An interesting phenomenon occurs when the source of a sound moves at a speed equal to or greater than the speed of sound. As an airplane's speed approaches that of sound,

the sound waves emitted by the plane's motor pile up in tremendous amplitudes in front of the plane and form a so-called sonic barrier that takes an extra burst of power to break through. To an observer on the ground, these sound waves constitute a shock wave that may be great enough to break windows in buildings. Once the plane exceeds the speed of sound, the pilot may no longer be able to hear the noise of his motor (if it is behind him) or the noise of the air as it rushes past him; he is traveling away from the source of the sound at a greater rate than the sound itself.

10-4 Source of Sound Waves

The frequency of audible sound is between approximately 20 cycles per second and 20,000 cycles per second. Some people may be able to detect sounds at somewhat lower or higher frequencies than these if the intensity of the sound and other conditions are favorable. The human ear, however, is not sensitive to excessively high frequencies and therefore cannot hear, for instance, the sounds emitted by a bat at frequencies of 50,000 cycles per second. Sound waves are longitudinal vibrations, and any oscillator, such as a tuning fork, that vibrates with a frequency in the audible range will produce sound waves that we can hear. There must, of course, be a suitable medium through which these waves can travel; sound waves will not travel through empty space because such space contains no substance that can be compressed and expanded by a wave passing through it.

Some of the most pleasing sounds are generated by musical instruments. A vibrating string gives off sound waves that may be made up of several frequencies. In Fig. 10-8 a string is vibrating at two frequencies—the fundamental and the second harmonic tone. The *fundamental,* or *first harmonic,* is produced by a standing wave that is twice as long as the string itself, and has the lowest frequency the string can produce. The *second harmonic* is produced by the string vibrating in two segments, and its frequency is twice that of the fundamental and one octave higher. The *third harmonic* is produced by the string vibrating in three segments, and so on. Actually, a violin string may vibrate simultaneously with several harmonic tones and produce a combination of frequencies that give the violin its own characteristic sound. Note that the vibrations in the string are *transverse* waves, whereas the string sends out a disturbance through the air in the form of *longitudinal* waves. Several types of instruments, including the violin, cello, and piano, are designed to produce variations in both wave length and harmonics by means of a vibrating string. The string may be set into vibration so that it will produce many harmonics simultaneously along with the fundamental vibration. The number and intensity of harmonics produced by a musical instrument determines the *quality* of its tone.

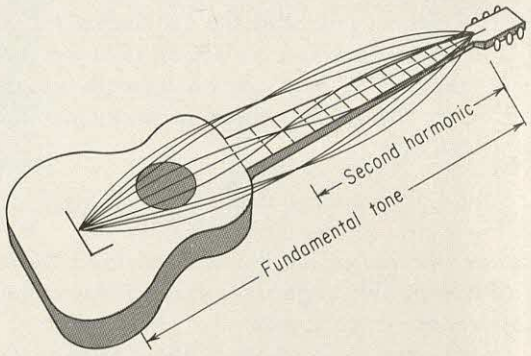


Fig. 10-8. The complex vibration of a string that shows both fundamental and second harmonic tones at the same time.

A very simple device for producing sound waves is a tuning fork (Fig. 10-9). If a vibrating tuning fork is placed at the open end of a tube of proper length, a series of standing waves is formed in the tube, and they intensify the tone by a process known as *resonance*. A number of musical instruments are constructed on this resonant-chamber principle. If a vibrating reed is used at one end of the tube, we have a clarinet, a saxophone, or a bassoon. The lips of the musician provide the vibrations for a cornet, a French horn, a trombone, and a tuba. In a flute the vibrations are initiated by an air whistle that creates pulsations of air in the upper end of the instrument. In all these musical instruments, the pitch is varied by changing the effective length of the tube and by thus changing the wave length of the standing waves that resonate in the tube. In the trombone the length of the tube is actually modified, and in others the tube's length is altered by a series of valves that open and close. In a pipe organ, the lengths of the tubes are fixed, and the different pitches are obtained by using a series of tubes that vary in length from 1½ inches to 32 feet.

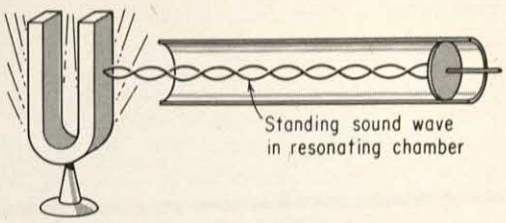


Fig. 10-9. The tuning fork is a source of relatively simple sound waves. The standing waves in the tube are an example of resonance.

The shapes of the waves produced by some of these instruments are shown in Fig. 10-10. Note that the fundamental tone or pitch of the notes is identical, because the frequency of the waves with the largest amplitude—the fundamental wave lengths—is identical. Superimposed upon the fundamental wave lengths are various sizes and shapes of shorter wave lengths that are produced by the harmonic vibrations. Each instrument has its characteristic group of harmonic tones, and, in much the same manner that we distinguish the sound of one instrument from another, we can distinguish them by the wave forms they produce in Fig. 10-10. The actual pitch of a musical note is changed by altering the wave length of the fundamental tone. In stringed instruments this is accomplished by changing the length and weight of the string; the highest notes are made by thin, short strings and the lowest notes by long, heavy strings. In wind instruments the length of the resonating chamber is changed.

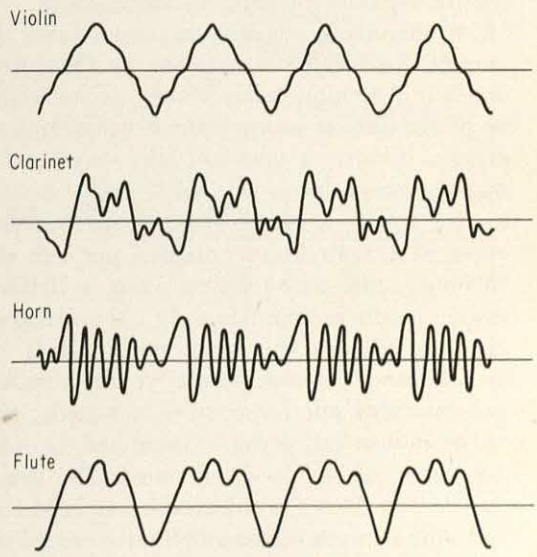


Fig. 10-10. The structure of the wave motions from various musical instruments.

A common, yet complex, source of sound waves is the human voice. It is probably the most versatile musical instrument in existence. The vibrating mechanism is the vocal chords in the larynx, and the resonating chamber is formed by the throat, mouth, and nose. Not only can the pitch of the vibrations be altered, but the size and shape of the resonating cavity can be modified so as to produce a wide range of tones. An interesting demonstration is sometimes conducted by inhaling and speaking with helium or a helium-oxygen mixture. Since sound travels faster in helium than in air, this makes the chambers of lungs, nose, and throat appear smaller, and the sound which is resonated has a higher pitch, which is surprising to listener and demonstrator alike.

Percussion instruments, such as drums and cymbals, pro-

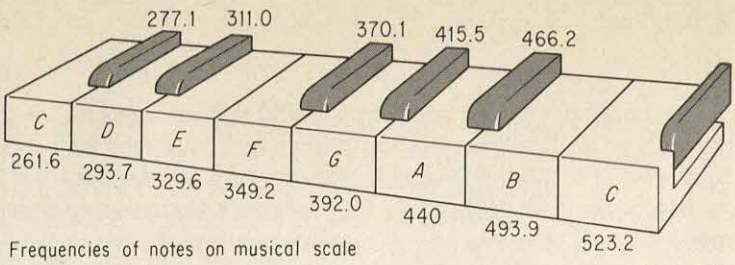


Fig. 10-11. The frequencies of the twelve notes in the middle octave on a piano.

duce sound waves that possess very little pitch, that is, they have no well-defined fundamental vibration frequency with accompanying harmonics. Technically, their sound waves should be defined as noise, although most bands and orchestras depend upon them for rhythm.

On the musical scale, the A note (orchestra A) is defined as that pitch which is produced by 440 vibrations per second. Twice this many vibrations, 880 cycles/sec, is A an octave higher, and 220 cycles/sec is A an octave lower. The musical scale is an octave divided into twelve equal-ratio intervals, as shown in Fig. 10-11. These intervals are half tones, and each one is represented by a white or a black key on a piano. The frequency of any note is 1.05946 times that of the note immediately below it.

An interesting sidelight on sound waves is the amount of energy that can be transmitted by them from one point to another. Although water waves, as we know, can transmit large amounts of energy, sound waves transmit only a little energy, because it does not take very much energy to produce ordinary levels of sound. Probably the noisiest thing in the world is an airplane. Both conventional and jet-powered aircraft convert about 1 per cent of their mechanical power into sound waves. Thus, a 10,000 horsepower jet engine produces approximately 100 horsepower of sound. To give you an idea how much noise that is, it has been estimated that 15 million people, yelling simultaneously, would generate only one horsepower of sound.

The human ear, perhaps unfortunately in a jet age, is very perceptive of low levels of sound. The slightest sound that the ear can detect is one whose amplitude is only 10^{-8} cm, and sound waves whose amplitudes exceed one or two millimeters in magnitude are painful to the ear. Riveting machines, sirens, explosions, thunder, and low-flying aircraft produce sounds on the threshold of pain, and the amplitude of such sounds may be great enough so that they can be felt.

The sound waves we have been studying in this section are longitudinal waves and are transmitted through some kind of medium. In the next section we turn our attention to electromagnetic waves which are transverse waves, needing no material for their transmission.

10-5 Sources of Light Waves

Light waves are produced by an oscillating source that is closely associated with the electrical and magnetic nature

of matter. Before considering the actual sources of light waves, however, let us turn our attention to a theory concerning electrical and magnetic forces that eventually led science into a very fruitful search into the structure of light. In 1856, James Clerk Maxwell proposed a theory based on his knowledge of electrical and magnetic behavior. He reasoned that if a wave consisting of electrical and magnetic forces could be generated it would travel with the speed of light. Light itself, he speculated, is probably an electromagnetic wave of much shorter wave length than the wave he suggested in his theory. With the information about electricity and magnetism that we gathered in the last chapter, let us examine Maxwell's theory.

In order to visualize Maxwell's electromagnetic wave, look at the diagrams in Fig. 10-12. In the tower *AB*, a single electron oscillates up and down the full height of the tower. When this electron is at *A*, a similar but positive charge will be at the bottom, *B*. An electron at *C* will be repelled by the charge at *A*, attracted by the charge at *B*, and the resultant force on *C* will be downward. A similar, but less intensive, effect will be noted on an electron at *D* and *E*. When the electron is at the bottom of the tower, the resultant force on the electrons at *C* and *D* and *E* will be upward. As the electron oscillates up and down the tower, therefore, it will exert an upward and downward force on the three electrons, as well as on all other electrons that come under its influence.

In addition to this electrostatic effect, you will recall that a moving electron also generates a magnetic field (Fig. 10-12b). Hence, as the electron moves downward in the tower, magnetic lines of force will be formed that will be horizontal, or at right angles to the resultant electrostatic force on the electrons *C*, *D*, and *E*. In other words, the electric and magnetic force vectors produced by an oscillating electron are at right angles to each other. Note that at *C*, *D*, and *E* the magnetic field acts horizontally at the same time that the electrical field acts vertically. Recalling the rules of electromagnetic induction in the last chapter, we see that the magnetic field affects the electrons at *C*, *D*, and *E* in the same manner that the electrical field does. For instance, when the electron in the tower moves downward from *A* to *B*, the electrical force on an electron at *C* is downward. As the electron moves downward in the tower the magnetic field, expanding outward and in a direction predicted by the left-hand rule (Fig. 9-12), also exerts a downward force upon the electron at *C*. Thus, the oscillating motion of the electron in the tower is transmitted to the electrons at *C*, *D*, and *E* by electrical and magnetic forces, causing these electrons to oscillate up and down at the same frequency as the one in the tower.

From this information, Maxwell concluded that if an electron could be made to oscillate, it would simultaneously emit electrical and magnetic forces that would affect charges

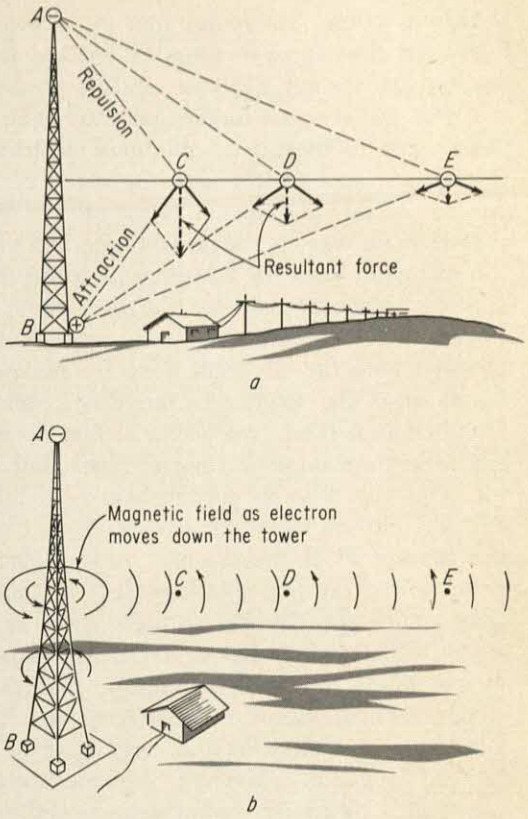


Fig. 10-12. The creation of electrical and magnetic fields by an oscillating electron.

at distant points. He found that there was an interval between the time these forces were emitted and the time they reached the distant electrons, and he predicted and proved that this lag is equal to the speed of light. An electromagnetic wave, in short, is a combined electrical and magnetic disturbance that travels with the speed of light. The wave concept of this disturbance is pictured in Fig. 10-13. Here, the electrical wave is shown vibrating in a vertical plane and the magnetic wave in a horizontal plane. These waves are emitted simultaneously by the oscillating electron in the tower, but note that the magnetic wave is out of phase (out of step) with the electrical wave by one-quarter of a wave length near the tower. As the waves move outward, the amplitudes of these two waves at first decrease rapidly because they are somewhat out of phase, but at a greater distance these two waves get in phase again, although they are still at right angles to each other. This, then, is essentially the concept of electromagnetic waves—electrical and magnetic waves traveling simultaneously as transverse waves at right angles to each other with the speed of light.

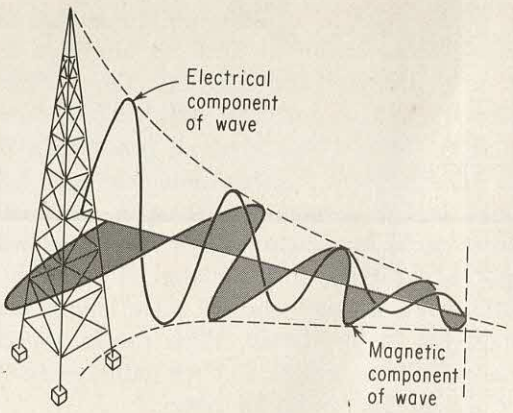


Fig. 10-13. An electromagnetic wave as the combination of an electrical and a magnetic disturbance.

In 1888, Heinrich Hertz, a German physicist, succeeded in generating an electrical impulse that gave off sufficient electromagnetic waves to be detected at a distance of a few hundred feet. Following the early work of Hertz, a young Italian, Guglielmo Marconi, became interested in these waves as a means of communication and sent out a signal from a wire stretched high into the air. He gradually increased the power of his waves, improved the method of detecting them at greater distances, and by 1900 had succeeded in sending a signal 200 miles. From these crude beginnings came our vast radio communication systems of today. The tower in Fig. 10-12 is a radio transmitting tower, and the oscillations of the electrons in the tower cause similar oscillations in the aerial of our radios many miles away. At the broadcasting station, audible sounds are converted into electrical energy that causes electrons to oscillate up and down the transmitting tower. Part of the energy of the electromagnetic waves sent out by the oscillating electrons reaches our radio receiver, causing electrons in it to oscillate in the same pattern as did those in the transmitting tower. In the receiver, the electron oscillations are converted back into audible sounds. The details of the electronic circuit involved here need not concern us. Our primary interest is in the fact that audible sounds can be converted into electromagnetic waves, either through modifying the frequency of the waves (frequency modulation or FM) or the amplitude of them (amplitude modulation or AM), and the fact that the electromagnetic waves can then be reconverted into sound waves.

The source of all electromagnetic waves is an oscillating electrical charge. For example, as we have already seen, radio waves, which are typical electromagnetic waves, are generated by electrons oscillating up and down in a radio

transmitting tower. The wave length of radio waves usually determines the height of the tower, which is usually one-half or one-fourth the wave length of the waves being generated. Radar uses a wave which is of shorter length than an ordinary radio wave. Ordinary radio waves vary from one meter to about one hundred meters in wave length, while radar waves range from one to one hundred centimeters in wave length. Hence, these waves must be generated by electrons that oscillate at a much higher frequency than those oscillating up and down a radio tower. Radar waves are generated in a fashion somewhat different from radio waves. In a so-called magnetron tube, surges of electrons are fed into the tube billions of times each second, and these surges, which resonate in a series of cavities in the tube, generate an electromagnetic wave that is literally piped through a wave guide (a hollow, rectangular brass tube) to the antenna. Although the details of the magnetron tube are somewhat complex, the principle of its operation is based on the idea that electrons, in oscillating at a frequency of 30 billion times a second, send out electromagnetic waves whose wave length is one centimeter.

Another form of electromagnetic wave is ordinary light. Here again, the source of the waves is oscillating electrons, although they oscillate at much higher frequencies than any encountered in radio and radar waves. Wave lengths of visible light vary from 7500 Angstrom units (an Angstrom, A, is 10^{-8} cm) for a red light down to 4000 Angstroms for violet light. These waves are produced by electrons that oscillate or vibrate within the atoms of the substance that emits light. In the ordinary light bulb, the little spiral of tungsten wire filament is heated by electrical energy, and part of this energy is converted to visible light (electromagnetic waves) as a result of the vigorous vibrations of the electrons in the tungsten atoms. Any solid, such as iron, carbon, and even quartz, when heated to glowing or incandescent temperatures, gives off light that is an electromagnetic wave.

Light waves are also generated by oscillating electrons in an atom when the atom is energized with electrical energy. The light emitted by ordinary neon signs, mercury vapor lamps, and sodium vapor lamps is the result of electrons vibrating through very short distances and with high frequencies in the range of 6×10^{17} times per second. We shall find in a later chapter that a study of the light given off by oscillating electrons in various kinds of matter reveals much about the nature of matter.

10-6 Reflection of Waves

All waves, except the shortest electromagnetic ones, may be reflected by barriers that stand in their path. You will recall that a standing wave in a vibrating rope or in a train of

sound waves results from the waves being reflected back over the path from which they came. To take another common example, light waves are reflected from a mirror, and if the angles at which the light approaches and leaves a mirror are carefully measured, we find that they are equal. Therefore, we say that the angle of incidence is equal to the angle of reflection (Fig. 10-14). The image we see in a mirror is called a *virtual image,* because it does not exist where it appears to be. This displacement can be dramatically shown by a simple experiment in which a candle is "burned" in a glass of water (Fig. 10-15). We see a *real image* when the reproduction of the object is where it appears to be, as in the case of a picture projected on a movie screen.

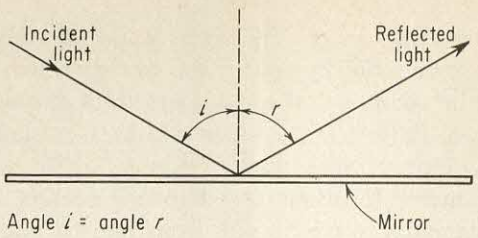


Fig. 10-14. The reflection of light waves from a flat mirror surface.

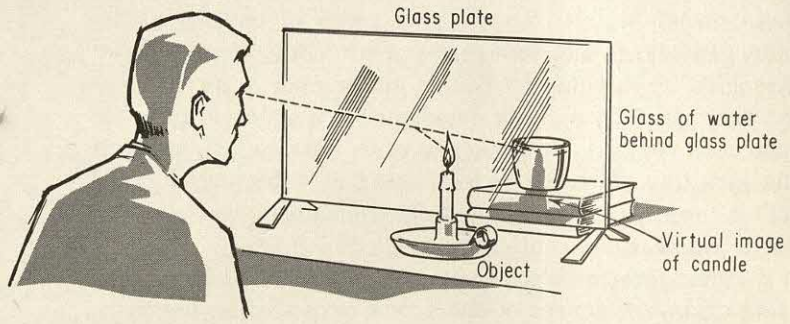


Fig. 10-15. A simple experiment showing a virtual image.

Other kinds of electromagnetic waves besides light can be reflected. Ordinary (AM) radio waves, for example, are known to travel greater distances on the earth's surface at night than in the daytime. The explanation lies in the existence of a special layer of gas some 50 to 150 miles above the earth that reflects the waves back to the earth at night. In the daytime, this layer apparently rises or vanishes. The layer of gas is composed of charged particles or ions that have been energized possibly by the earth's magnetic field. In honor of its discoverers, this phenomenon is called the Kennelly-Heavyside layer; it was originally proposed in 1902 and was verified experimentally in 1925. Radar waves, though not reflected by the Kennelly-Heavyside layer, are reflected by almost all kinds of solid objects, and are consequently used primarily for the detection of such objects. They are sent out and then reflected back to the transmitter from, say, a ship at sea. The time lapse between the transmission of the wave and its reception back at its source gives us the distance from the transmitter to the ship.

At some time or other we all have experienced the reflection of sound waves in the form of an echo. Sound waves are reflected whenever they strike a boundary between two media of different density. The walls of a room, the edge of a forest, the side of a hill, and the bottom of the ocean are common examples of boundaries that act as a "mirror" to sound waves. In order for the sound waves to return to their source, the barrier must be at right angles to the direction in which they originally traveled.

The reflection of sound waves is used extensively in pros-

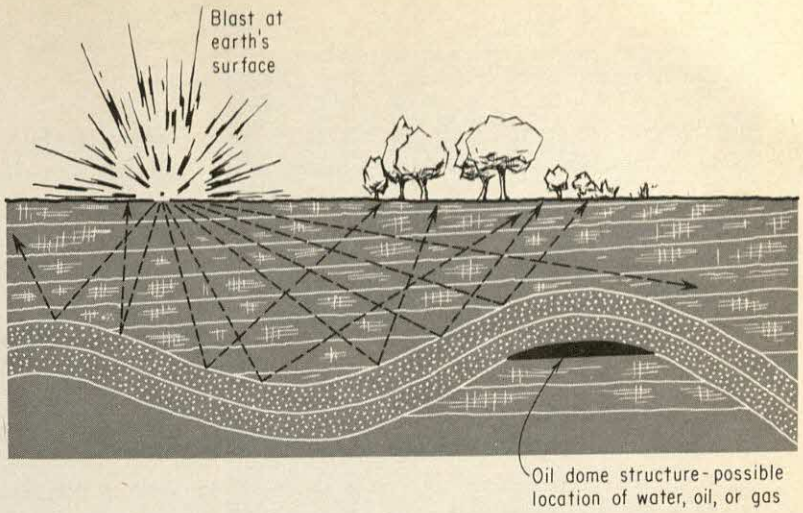


Fig. 10-16. Prospecting for mineral resources with sound waves.

pecting for certain mineral deposits, especially for oil. Oil, and also gas and water, often collect under anticlinal folds that are capped by an impermeable layer of rock, which is frequently shale. The shape of the oil dome may be detected by "reflection shooting." A charge of dynamite is set off at the surface (Fig. 10-16), and as the sound waves travel outward in all directions, they are reflected from the layers of rocks that lie below the surface. Seismic detectors at various points on the surface register the time lag in the waves that are reflected, and, although analyzing this information requires considerable training and experience, the shape of the buried rock strata can be determined from such information. Several oil fields of major size have been discovered in this way, and this method continues to be widely used for petroleum prospecting in many areas.

The depth of the ocean is now easily measured by a process known as "sonic sounding." Instead of lowering a weight on a wire or rope, ships today are equipped with a device on their bottoms for emitting sound waves that are reflected from the ocean floor. Although it is used primarily as a safety device, ships employ this mechanism to keep a continuous record of the ocean's depth while they are under way. The contours of the ocean floor have been well charted by this means, especially the topography beneath the shipping lanes. To show how simple this system is, let us calculate the ocean's depth at a point where it requires 8 seconds for the sound wave to travel from the ship to the sea bottom and back. From Table 10.1 we know that the speed of sound in water is 4700 ft/sec. Since the sound travels downward for 4 seconds, it descends a distance of 4 sec $\times$ 4700 ft/sec, or 18,800 feet, which is the depth of the ocean at that point.

10-7 Refraction of Waves

Whenever a wave travels from a more dense to a less dense medium, or vice versa, it is bent from its original direction. The primary reason for this bending effect is the difference in the velocity of waves in various media. For example, sound

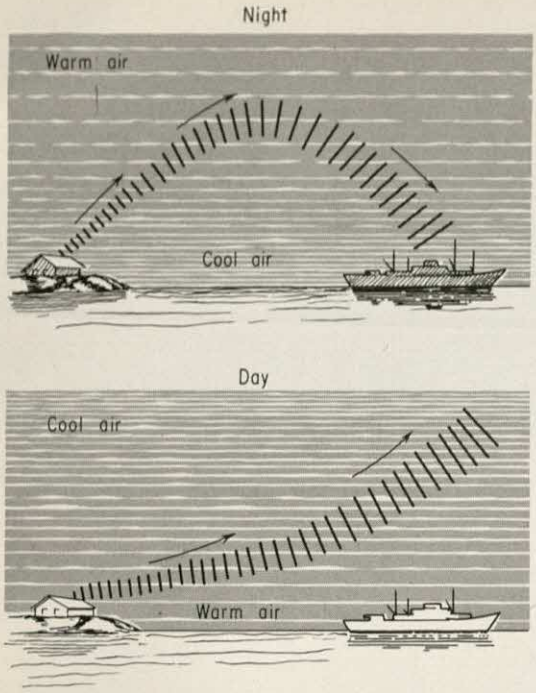


Fig. 10-17. Since sound waves travel faster in warm air than in cool air, they are refracted by air of non-uniform temperature.

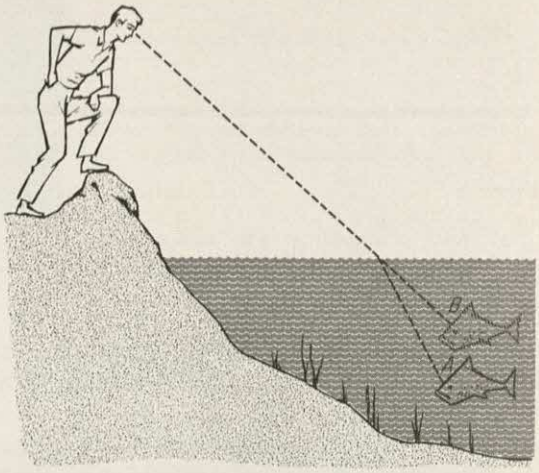


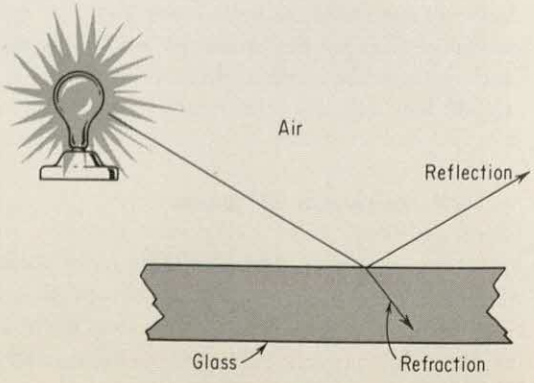
Fig. 10-18. The refraction of light waves as they pass from a more dense medium (water) to a less dense medium (air).

waves travel faster in warm air than in cool air. We know that sound carries great distances at night over water, but in the daytime the same sound travels only short distances. The reason for this difference is explained by Fig. 10-17. At night the air near the water is cooler than that farther above the water, and the sound waves that normally would continue to travel upward and be lost to hearing are refracted downward. In the daytime cool air is on top of the warm air, and sound waves are bent upward. This bending of waves as they pass from one medium into another where their velocity is different is called *refraction*.

Light waves, too, travel at different velocities in different substances. We are so accustomed to seeing light waves travel in straight lines that we are often deceived by the refraction of light. A fish in clear water always appears at a shallower depth than he actually is (Fig. 10-18). Light traveling from the fish at *A* is bent or refracted at the surface of the water, so that the fish looks to us to be at *B*. The velocity of light in water is only three-quarters of its velocity in air. In glass, the velocity of light is even lower—two-thirds of its velocity in air.

When light waves strike a smooth glass surface, part of the light is reflected and the rest is refracted (Fig. 10-19). We will see later that the capacity of glass to refract light is utilized in many optical instruments. The mysterious appearances of mirages along flat stretches of highway, especially in desert country, are produced by the refraction of light. Although the velocity of light does not vary greatly between warm and cool air, it does vary enough to cause us to see

Fig. 10-19. Refraction and reflection of light waves as they strike a boundary between substances of different density.



what appears to be water in the distance. And what is it that we actually see? It is the refraction of light from the sky.

10-8 Interference of Waves

All waves are vibrational disturbances that periodically go through crest and trough formations as they pass through space or other transmitting media. The wave that is sent along a stretched rope, for instance, shows alternate crests and troughs as the wave passes through the rope. We may think of the amplitude of the crest as a certain amount of displacement upward, and the trough as an equal amount of displacement downward. In other words, the displacement of the crest is equal but opposite in effect to the displacement of the troughs. Now let us send two waves of equal wave length and amplitude along the rope simultaneously, but arranged so that the crests of one wave occur exactly where the troughs of the second wave occur. These two waves will exactly cancel each other because the upward displacement of a crest will be counteracted by the downward displacement of a trough, and the rope will remain motionless. When such a situation occurs we say that there is *interference* of the waves, destructive interference, as it were, in this case.

The interference of light waves may be readily examined by an experiment similar to the reflection of light shown in Fig. 10-14. Let us use a very thin mirror with a glass about 6000 A thick. (The pages of this book are about 400,000 A thick.) As light strikes the surface of this thin piece of glass some of it is reflected. The rest of it passes into the glass, and at the back side of the glass, the remaining light is reflected. In Fig. 10-20a the crests of the light waves are shown as a series of three parallel lines that are perpendicular to the direction of the wave motion; the troughs are at the blank areas between the crests. In the illustration we see that the crests of the waves reflected from the bottom surface of the glass coincide with the crests of the waves reflected from the top surface. But in Fig. 10-20b the angle of incidence is such that the crests of the reflected waves do not coincide but cancel each other, and we can conclude that at this particular angle of incidence there can be no reflection of the light waves illustrated in the figure.

If the light used in the above experiment is ordinary white light—composed of wave lengths from 7500 A to 4000 A—we can expect that at a given angle of incidence certain wave lengths will undergo destructive interference, or cancel, while the waves of longer or shorter wave lengths, reflected by both the front and back surfaces, will constructively interfere and reflection of light will occur. Thus, the white light should now be separated into colors because certain wave lengths are destroyed. The color we see in the light reflected by soap bubbles and by thin layers of oil on water are the result of such interference.

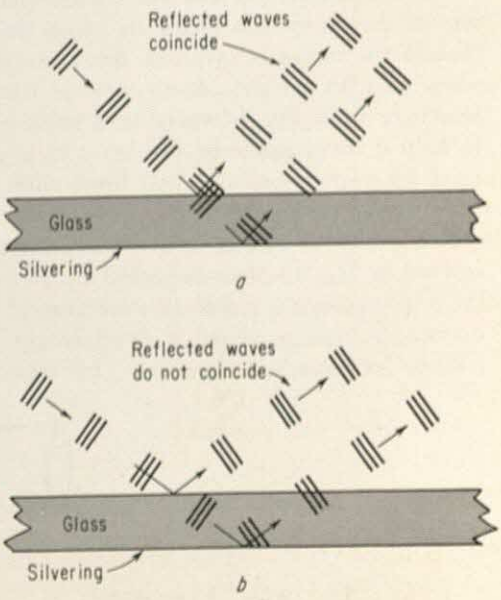


Fig. 10-20. Interference of light waves reflected from a very thin piece of glass.

Interference of sound waves is noticeable when two musical notes of almost the same freqency are sounded. For instance, if two whistles, one with a frequency of 240 cycles per second and the other with 241 cycles per second are blown simultaneously, there will be a slight wavering, or *beat*, each second. When crests match crests, the two whistles sound as one, but into each second there must be crowded an extra wave length for the second whistle, so part of the time the crests from one whistle will coincide with the troughs of the other and actually cancel the sound as the result of the interference. This effect is purposely achieved in the pipe organ where two pipes are tuned slightly apart. The resulting beat notes produce a wavering or a tremulous effect that is essentially the interference of sound waves.

10-9 Optical Instruments

Optical instruments are designed to use light waves in various manners. In this section we will examine three of the most important such instruments—the camera, the telescope, and the spectroscope. The camera and the telescope rely primarily upon a lens, which is a glass disc that refracts light. A lens is either thicker at the center than at the edge (convergent), or thicker at the edge than at the center (divergent). Cross sections of these two types of lenses and their effect upon light are shown in Fig. 10-21. These lenses are called convergent and divergent because they make the light rays that pass through them either converge or diverge. In a camera, a convergent lens focuses the image of the object being photographed upon a sensitive film at the rear of the camera (Fig. 10-22). Note that the image produced on the film is inverted.

The eye, which in many ways is comparable to a camera, has its own lens that forms an image on the retina. Even though the image is inverted, our sense of perception provides us with a right-side-up view of the world. The eye's lens may be faulty, however, and require an auxiliary lens to help it focus properly. Glasses with a divergent lens are used for nearsighted eyes and those with a convergent lens are used for farsighted eyes.

A telescope is similar to a camera, except that the image formed in Fig. 10-23 is magnified by an eyepiece lens. This kind of telescope is called an astronomical telescope because its inverted image is not a disadvantage in making astronomical observations. However, for terrestrial observation,

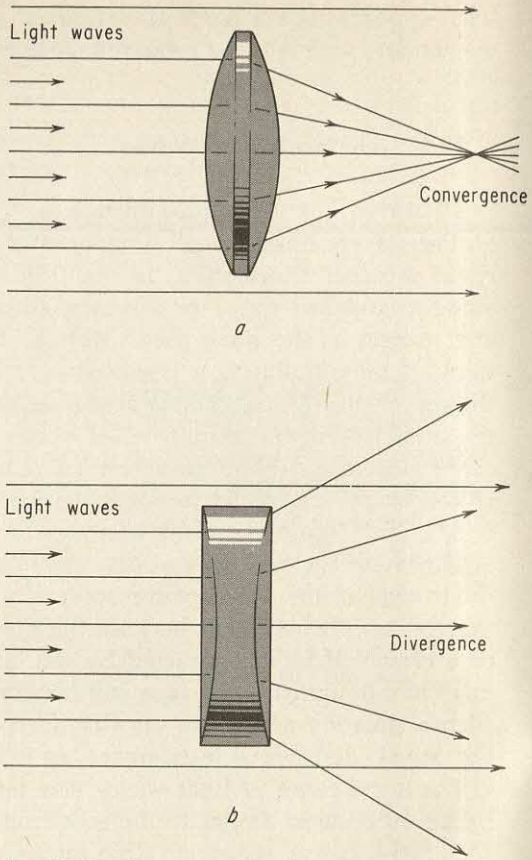


Fig. 10-21. Cross-sections of a convergent and a divergent lens.

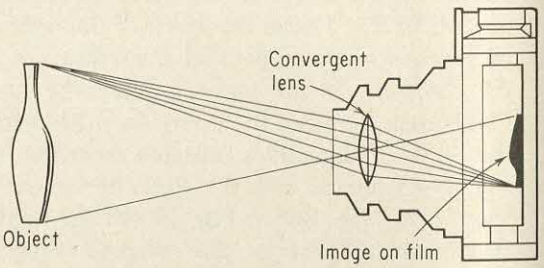


Fig. 10-22. A camera.

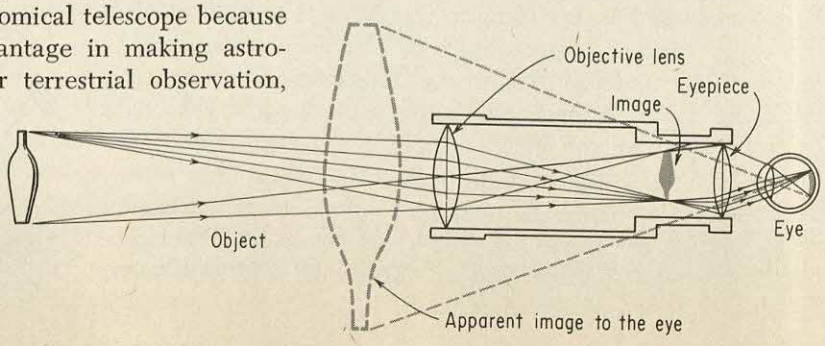


Fig. 10-23. A telescope.

where we prefer to see an image right side up, an erecting system is needed to right the inverted image. This may be accomplished by re-inverting the image with an additional set of lenses or with a set of prisms such as those that are used in a prism-type binocular.

Many optical instruments make use of the *prism*, which is a wedge-shaped piece of transparent glass or other material (Fig. 10-24). Since light is refracted whenever it passes through a boundary between two media of different density, the light passing through the wedge is bent from its original path. From the earliest times, it has been known that the light that passes through transparent crystals and jewels is colored, but it was Newton who first showed that these colors are the result of the original light being separated into its various color components. Consequently, sunlight, which appears as a white light, will yield all the colors of the rainbow when it is passed through a glass prism.

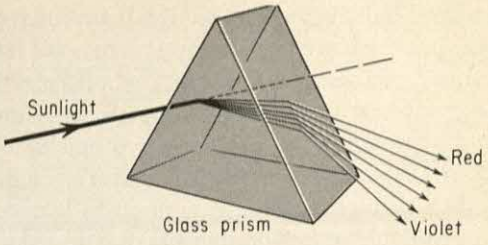


Fig. 10-24. Light separated into colors by a prism.

In Fig. 10-24 we can see how sunlight is separated into colors. All the light is refracted, but the shorter wave lengths (blue and violet) are refracted more than the longer wave lengths (red and orange). The spread of these colors is called a *spectrum*, which is simply a distribution of wave lengths. In the study of spectra, the scientist uses a *spectroscope*, an instrument that separates the colors of light by means of a prism. We will find in later chapters that studies of light waves with the spectroscope have resulted in some of the most exciting discoveries in science.

10-10 Color

A world without color would be drab indeed. We would have no blue sky, no brightly colored sunsets, no green grass, and no blondes, brunettes, or redheads! Color vision, then, is perhaps the most valuable sense we have. Color may be produced by several kinds of sources. A hot glowing body, such as the sun or the filament in an incandescent lamp, gives off all colors of light. We speak of these mixtures of colors as white light, and a prism will spread this light into a continuous spectrum—a solid bank of colors from red through orange, yellow, green, blue, indigo, and violet. A glowing body at very high temperatures will be bluish white, that is, the colors on the violet end of the spectrum will be more intense than those on the red end. At temperatures only sufficient to cause a body to glow, the color will be reddish (Fig. 12-6). Astronomers use this relation between color and temperature to determine the temperature of the stars. Even with the unaided eye, you can note considerable differences in the colors of stars.

Color is also produced by gases at relatively low temperature when they are suitably excited by electrical energy. If all the air is pumped out of a glass tube and then a small amount of sodium vapor or neon is admitted, the tube will

give off light when an electrical current is passed through it because the electrical energy is converted into light. A sodium vapor lamp and a neon sign give off light that ranges only a short distance on the spectrum. If we view a sodium vapor lamp through a spectroscope, we find that its yellow-orange color is produced by only two wave lengths of the emitted light; on the spectrum, it appears as two very closely spaced lines in the yellow-orange band. The color of a neon lamp, viewed through a spectroscope, is composed of several discrete lines in the red-orange portion of the spectrum. The light given off by these excited gases is called a *line spectrum* because only certain colors show up in the form of distinct color lines in the spectroscope. No two gases produce identical line spectra, that is, each gas emits its own peculiar set of colored lines when it is excited by an electrical current; each element, so to speak, has its own set of fingerprints. A study of the various colors emitted by substances such as iron, sodium, neon, and so on led to an important discovery in atomic structure, the details of which will be discussed in Chapter 12.

A red flower, a blue tie, and a green dress are obviously not glowing bodies. They get their color from their surfaces and the characteristic light waves they reflect. These colors are called *pigment colors*. The white light—which, as you will recall, is composed of all colors—that strikes a red flower is entirely absorbed, except for the red light, which is reflected and gives the flower its color. When we view variously colored objects under colored light, we get some interesting effects. A blue tie in the light from an orange-red neon lamp will appear black simply because there is so little blue light in the neon light that the blue tie can reflect. A green dress appears black in the yellow-orange light from a sodium vapor lamp. For pigment colors to appear in their natural colors, they must be viewed in white light, or in the color of light they are capable of reflecting.

Structural color is the color produced by certain physical arrangements that sort out various colors. The sky is blue because the dust and other particles in the atmosphere are capable of scattering blue light more than other colors. The blue feathers on birds contain no blue pigment or dye; they happen to be constructed in such a way that they scatter only blue light. The feathers of bluebirds have been carefully analyzed and found to contain no blue coloring matter whatsoever. A green bird has, in addition to its ability to scatter blue light, a yellow pigment in its feathers. The combined effect of the yellow pigment and the scattering of blue light accounts for its green appearance.

10-11 Rainbows

One of nature's most spectacular shows of color is the *rainbow*. Whenever the sun shines into a raindrop, a rain-

bow is visible, because each raindrop acts as a tiny prism and separates the sun's light into its spectrum (Fig. 10-25). Only at a particular angle between the sun, the raindrop, and between the observer does the raindrop separate out and reflect a particular color; however, at other angles, it is either transparent to the light or it reflects it. Note that the light that forms the rainbow enters the top of the drop and is refracted. It travels across the drop to *B*, is reflected to *C*, where it is refracted again into a *primary rainbow*. The colors are the result of the fact that the shorter wave lengths are refracted more than the longer wave lengths. Under certain conditions, it is possible to see a double rainbow, and the weaker of the two, the *secondary rainbow*, is caused by sunlight entering the bottom of the raindrop, being reflected at two points on its inner surface, and then refracted out through its upper part. Since the angle here between the incident and refracted light is slightly larger than the same angle in the primary rainbow, the secondary rainbow will appear above the primary rainbow.

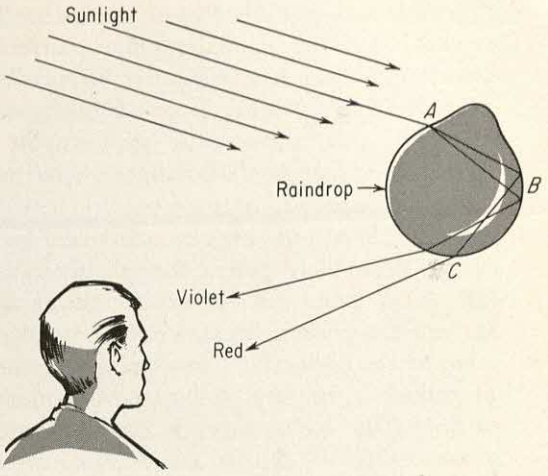


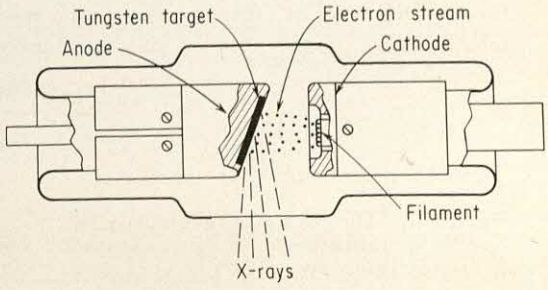
Fig. 10-25. A raindrop serves as a prism when it forms a rainbow.

10-12 X Rays

X rays are a unique type of wave motion that deserves special consideration. We are fortunate to live in a world that can make use of X rays—to examine the soundness of the axle of a car, to help a surgeon find a broken bone, to serve as an indispensable tool for the scientist in his study of matter, to mention just three ways in which X rays assist in the advance of our civilization. The discovery of X rays occurred quite by accident. In 1895, Wilhelm Roentgen, while studying the fluorescence, or pale greenish glow, that occurs in a cathode, or Crookes tube, noted that under certain conditions barium salts at considerable distance from the tube also fluoresced. The tube emitted a strange kind of radiation that went through the glass walls of the Crookes tube and even through black paper, wood, and many other substances that are ordinarily opaque to light. Roentgen called this penetrating radiation X rays, the letter "X" standing for the unknown.

A series of experiments soon indicated that these X rays came from the point where the cathode rays struck the glass wall of the tube. When the cathode rays were focused on a metal target, a greater intensity of X rays was emitted. A modern X-ray tube (Fig. 10-26) bears little resemblance to the early model, although all X-ray tubes operate on a principle that will be described in more detail in Chapter 12. At this point, we can merely say that some of the energy of the stream of electrons causes electrons within the atom to oscillate at extremely high frequency—in the order of 10^{18} vibrations per second; X rays are consequently another kind of electromagnetic wave. Roentgen's discovery was a significant milestone in physical science, and because the study

Fig. 10-26. A modern X-ray tube.



of X rays and their properties has led to the formulation of several important principles and concepts, X rays have been described as "the key to modern science."

The wave length of ordinary visible light is about 10^{-4} cm, whereas X rays travel with wave lengths of approximately 10^{-8} cm, which is 10,000 times shorter. Ordinary light is stopped, absorbed, or reflected by most substances; even window glass stops certain colors of light, particularly the colors on the violet end of the spectrum. As the frequency of light is increased and the wave length becomes shorter, solid materials in general become more transparent to electromagnetic waves. When the wave length becomes about the same magnitude as the size of the atoms themselves, as is the case of X rays, the waves travel through the material quite readily. Waves with still shorter wave length are even more penetrating than X rays.

At this point, we may profitably summarize the information we have about the wave lengths of the various kinds of electromagnetic waves. In Fig. 10-27 a complete electromagnetic spectrum from the shortest to the longest wave lengths is shown. On the left are the shortest waves, gamma rays (about which we learned in Chapter 8), which are the most penetrating because they have the shortest wave length. Next come X rays and then ultraviolet waves. In the ultraviolet range is a group of electromagnetic waves that are less penetrating but have sufficient energy to cause sunburn and the destruction of living tissue. Note that the visible portion spans only a small fraction of the whole electromagnetic spectrum, for we cannot directly sense any of the shorter or longer wave lengths. Heat energy is transmitted as an electromagnetic wave in the infrared region, and the direct heat that comes to us from the sun travels as radiant energy in this region of the spectrum. At wave lengths longer than heat waves come radar and radio waves. Even though there is a tremendous range in wave lengths—from 10^{-12} cm to more than a thousand meters—all electromagnetic waves travel with the same velocity, 3×10^{10} cm/sec, or approximately 186,000 mi/sec, which is the speed of light.

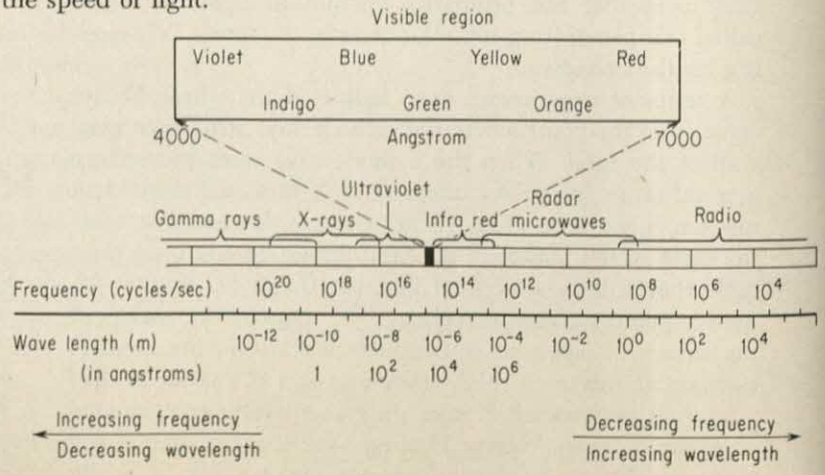


Fig. 10-27. The complete electromagnetic spectrum.

10-13 Wave Motion in Natural Processes

In Chapter 6 we saw how the shock of an earthquake sets up wave motions within the earth as well as on its surface. We can now better understand the difference between the primary, or longitudinal, earthquake waves and the secondary, or transverse, waves. Primary waves travel approximately twice as fast as secondary waves, and we saw that a record of these two waves, noted on a seismograph (Fig. 6-6) at a point some distance from the earthquake, reveals the distance the waves have traveled.

With our introduction to reflection and refraction in this chapter, we can now examine in more detail the behavior of waves as they travel through the earth and the information they give us about the earth's interior structure. If a disturbance, such as an earthquake, originates at point *A* in Fig. 10-28 and is subsequently recorded on seismographs at points *B*, *C*, *D*, and *E*, its rate of travel can be determined. Note that the wave sent from *A* to *B* may travel near the surface or far below it. At the surface the primary wave travels at varying speeds because of the variations in crustal composition—granite, limestone, and so on. But in the region directly beneath the crust, we find that primary waves travel about 8 mi/sec and secondary waves about 3 mi/sec. At greater depths, these waves not only move more rapidly, but they bend or refract as the result of the greater density of the material in the lower part of the mantle.

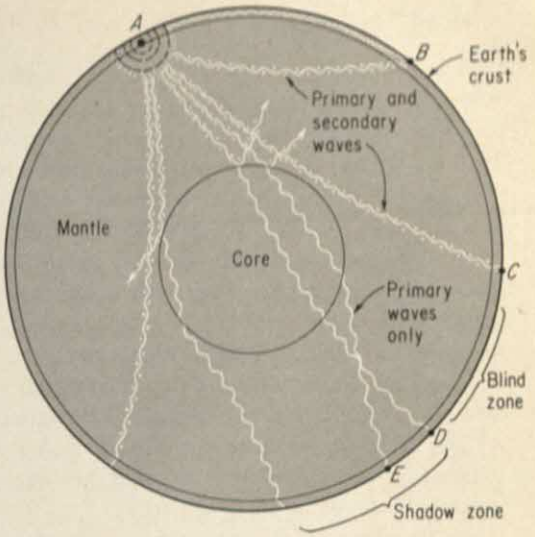


Fig. 10-28. Wave motion in the study of the structure of the earth.

As we saw in Chapter 6, there is a discontinuity at the bottom of the mantle, some 1775 miles from the earth's surface. Here, the primary waves are refracted and reflected, but the secondary waves are only reflected. Thus, a seismograph on the side of the earth opposite the original disturbance would fail to detect any secondary waves because it would be located in the "shadow" of the core. It is this "shadow" that leads us to the idea that the core, or at least its outer part, must have the characteristics of a liquid because a liquid is capable of transmitting longitudinal but not transverse waves. Primary waves travel through the core, but are refracted by the core, which serves as a convergent lens and bends these waves, concentrating them at the back side of the core. Consequently, there will be a shadow zone on the surface opposite the original disturbance where primary waves will be detected and an adjacent blind zone that receives neither primary nor secondary waves.

Our picture of the interior of the earth is based on indirect evidence and several broad assumptions. Among scientists there is considerable debate about how correctly our laws and theories fit the conditions inside the earth where pressures are millions of times greater than at the surface and temperatures are possibly as high as 5000°C—the temperature of the surface of an ordinary star. The earth's interior is truly a frontier we are just beginning to explore.

10-14 Summary

Wave motion is a fluctuating or oscillating motion by which energy is transmitted from one point to another. The wave may use a substance as the transmitting medium as in the case of sounds or earthquakes, or the waves may be transmitted in the absence of matter as in the case of electromagnetic waves. Particles of any transmitting material may move back and forth parallel to the wave itself (longitudinal wave) or back and forth at right angles to the wave (transverse wave). Waves on the surface of a liquid are complex, and particles of material move in a circular motion.

The wave length is the distance from one crest to the next, and the amplitude is the distance of maximum displacement of the particles as they move back and forth in the wave motion. The frequency is the number of crests that travel past a given point in a second.

If the source of a wave moves toward or away from the observer, the Doppler effect is noted. The frequency of a wave is increased in the first case, decreased in the second.

All waves are produced by an oscillator. Audible sound waves are produced by an oscillator that vibrates between 20 and 20,000 times per second, setting up longitudinal waves.

Electromagnetic waves are produced by the vibrations or oscillations of electrons. These waves are the result of an electrical wave and a magnetic wave (both are transverse waves) traveling at right angles to each other, and they move at the rate of 3×10^{10} cm/sec or about 186,000 mi/sec. The wave lengths of electromagnetic waves vary from 10^{-12} cm to 10^{6} cm.

Waves are reflected whenever they pass from a medium of one density to a medium of another density, and at this boundary they may also be refracted.

The camera, telescope, and spectroscope are instruments, indispensable to the scientist, that are based on the optical properties of wave motion.

Color is produced by hot, glowing objects, by excited atoms, by reflection from pigmented bodies, and by the selective absorption, scattering, or diffraction of white light.

The study of the shock waves produced by earthquakes gives us a picture of the structure of the interior of the earth.

EXERCISES

A. KEY TERMS TO REMEMBER

amplitude
convergent lens
crest
Doppler effect
electromagnetic wave
frequency
longitudinal waves
nodes
octave
prism
radio waves
rainbow
reflection
refraction
sonic barrier
spectrum
standing wave
wave length
X ray

B. QUESTIONS ABOUT WAVE MOTION

1. Compare tranverse and longitudinal waves in their effect upon the medium through which they travel.

2. If a car traveling 70 mi/hr sounds its horn, will the pitch of the horn to another car traveling 70 mi/hr in the same direction be the same, higher, or lower, than the original?

3. What characteristics of a sound wave determine the pitch of the sound?

4. What differences in the sound waves would you find in comparing noise with a musical note?

5. Produce standing waves experimentally as in Fig. 10-2 and determine the effect of the rope's (a) tension, and (b) its mass on the number of waves produced.

6. Describe the oscillating source that produces sound waves in several kinds of musical instruments.

7. By means of the Doppler effect, explain why the color of a star that is moving away from us will appear shifted toward the red end of the spectrum.

8. After examining a 2 × 2 slide projector, draw a diagram of its optical system in a manner similar to Fig. 10-23.

9. Draw a diagram of the human eye showing why objects close to a farsighted person seem blurred. Using a colored pencil, on the same diagram show how a lens can correct this defect.

10. Describe the electrical and the magnetic forces that accompany an electromagnetic wave. Tell how these forces constitute the wave itself.

11. When an airplane exceeds the speed of sound, a strong shock wave is formed that may shatter windows. What is the source of the energy in this shock wave?

12. Why does the blue color of the sky change to yellow, orange, and red at sunset?

13. Draw a diagram similar to Fig. 10-25 to show how light entering a raindrop will produce a secondary rainbow.

14. If there were no atmosphere, what would be the color of the sky? What is the color of the sky to an astronaut who ascends to a height above the dust and atmospheric layers that ordinarily cause the sky to be blue?

15. Using three tuning forks of differing but known frequencies and a long glass tube which can be filled with water, determine the relationship between the length of the tube and the wave length of the sound with which it is resonant.

16. Make a list of the uses to which X rays have been put.

17. Approximately how thin should a piece of glass be in order to cause interference of X rays in a fashion similar to the interference of visible light shown in Fig. 10-20? Is it possible to make so thin a sheet?

18. Besides those waves specifically mentioned in the chapter, can you think of examples of other wave phenomena?

Eleven

Atomic Theory

At this point in our study of physical science, let us pause for a moment in order to gain a wider perspective. On the basis of our study so far, what do we know about a given object, say a marble? If we were to examine the marble we could specify several of its basic attributes. It has a certain mass, say ten grams or 0.01 kg. It is attracted by the earth's field of gravity with a force of 0.01 kg times 9.8 m/sec^2, or 0.098 newtons. Because of its mass, the marble possesses inertia, and because of its mass and velocity, it has momentum and kinetic energy. Its density is related to both its mass and its volume. The position of the marble determines its potential energy. If we deposit an electrical charge on the marble, then it may have an electrical potential, or voltage. Unless the marble is made of a substance such as iron it cannot be magnetized, even though it is placed in a magnetic field.

A marble is an object of ordinary size—we can easily see it and handle it, but in the natural world about us a marble is only about midway in size between the largest and the smallest objects that we have encountered up to now. The sun is 50 billion times larger than the marble, but the marble is 100 million times larger than an atom—the basic unit of all matter. It is about as difficult to comprehend the small size of the atom as it is to grasp the hugeness of the sun. We cannot see an atom, nor can we touch or handle it as a single entity. There are nearly 10^{24} atoms in our marble, a number so large that it has little meaning, and if we could place all these atoms touching each other in a row, they would form a continuous line around the world a million times.

Our present-day view of atoms is based on centuries of study that has stemmed from man's strong curiosity about the inner arrangement of the world around him. Ordinarily, in our study of physical science in this book, we have not traced the historical development of ideas and discoveries. We have occasionally referred to the name of the person who made the discovery, but usually we have not delved into the details of these early developments. In this chapter, however, in order to more fully appreciate the development of the modern concept of atoms, we shall make a brief survey of the history of man's attempt to explain the structure of matter in terms of its ultimate composition. Even though the idea of atoms is now generally accepted, the layman may more adequately understand the importance of the atomic theory in science if he follows the origin and growth of the idea.

11-1 Historical Theories of Atoms

If we turn philosopher for a moment, and think about the ultimate structure of matter, such as that in a piece of iron, we might speculate on what would happen when we divide and subdivide the iron into smaller and smaller bits. Eventually one of two things must happen: We can either divide the iron into increasingly smaller particles forever, or we suddenly encounter a particle that resists division. In the first case matter is a continuous, infinitely divisible material, and in the second case it is composed of tiny indivisible particles.

Just where and when these two opposing ideas were first proposed, and by whom, are not known. Since the word "atom" comes from the Greek word *atomos*, which means "indivisible," we might suppose that the Greeks were the first to conceive of atoms. Although it is probable that philosophers in Asia and Egypt considered this problem, too, it is generally believed that the debate on the ultimate structure of matter was first vigorously argued among the Greek philosophers during the fourth century B.C. Aristotle held that matter is composed of a formless *hyle* that is continuous and can be subdivided without limit. Against this idea, Leucippus of Miletos, a flourishing city on the coast of Asia Minor, and his pupil Democritus maintained that matter is indestructible and composed of indivisible particles called *atoms*. These atoms were assumed to be uncreatable, indestructible, infinite in number, and endowed with inherent shapes and forms that enable them to react with one another. In addition, they believed that these atoms had been in constant motion since the time of creation, and that collisions between them produced various substances, such as water, air, metals, and rocks. The Greek philosophical concept of atoms, however, did not survive the decline of Greece, and it is doubtful if it contributed directly to the formulation of our modern

atomic theory. Nevertheless, we should not view the early Greeks as being entirely ignorant and incapable of explaining the structure of matter. Rather, we should recall that in these early times man was merely beginning to explore the intimate details of nature. We can no more expect these early philosophers to have generated a full-fledged atomic theory than we can expect their successors, the Roman warriors, to have gone forth into battle girded with an atomic bomb.

The atomic theory of Leucippus thus lay dormant for two thousand years, with only an occasional insignificant revival. During the seventeenth century, however, Robert Boyle (an Irishman who became one of England's first modern scientists), Sir Isaac Newton, and other scientists rejuvenated the atomic concept even though they had no direct experimental proof for it. Boyle proposed that all substances are made up of *elements* that cannot be separated into any simpler substances by ordinary chemical processes. For instance, iron is an element because it cannot be separated into anything simpler than iron, but limestone is not because it can be separated into three elements—calcium, carbon, and oxygen.

11-2 The Chemical Revolution

For the most part, there was little experimental support for these historical theories, yet these philosophical approaches stimulated scientists to band together into societies where they discussed these ideas. These discussions led to tentative conclusions which could be tested in the laboratory, and experiments were increasingly used to prove and disprove ideas. Instead of endlessly debating the merits of a particular theory, scientists during the seventeenth century began to devise experiments to test their theories. It is in this period—sometimes called the scientific revolution—that the modern scientific era dawned. Laboratory investigation became recognized as an integral part of science, and by the end of the eighteenth century a series of brilliant experimental discoveries were made that opened the door to an understanding of the atomic nature of matter.

During this period—the latter part of the eighteenth century—Joseph Priestley, an English clergyman and amateur chemist, greatly improved the techniques known at that time for producing and studying gases. He designed laboratory apparatus by which he could confine gases, measure their volumes, and investigate their various properties, and he used this equipment to study some of the gaseous "chemical elements" that Boyle had predicted earlier. Priestley was among the first to prepare and examine such gases as oxygen, carbon dioxide, nitric oxide, and ammonia. He made oxygen by heating a sample of mercury oxide enclosed in a glass tube by focusing the sun's rays upon the sample with a large convergent lens. In his later years Priestley moved to

America, and this burning lens is preserved along with many other pieces of his equipment at his home in Northumberland, Pa.

The gas, oxygen, that Priestley prepared turned out to be an element because it could not be decomposed into anything simpler. By now about a score of elements had been isolated—gold, silver, mercury, iron, copper, lead, and nonmetallic substances such as sulfur and carbon. But there were several common substances, such as water and lime, that were believed to be elements, although they were later decomposed into simpler components. The composition of water was examined by Henry Cavendish, a lonely genius and one of the wealthiest men in England. He made a new gas—hydrogen—by adding an acid to a metal such as zinc. He studied this gas by exploding various mixtures of it with air, and, as a result of these noisy experiments, he discovered that about four-fifths of air is a relatively inactive component (later called nitrogen) and that one-fifth of it is an active component (oxygen) that enters into combustion with the hydrogen. After repeated explosions of hydrogen-air mixtures he discovered in the bottom of the explosion chamber a substance which he described as "a clear liquid that appeared to be ordinary water." Although the exact meaning of the term *chemical element* was not generally understood at this time, these experiments pointed up the big difference between elements and compounds, and it gradually became evident that elements, such as hydrogen and oxygen, are simple substances and that compounds, such as water, are combinations of elements.

One of the most brilliant men of all science, Antoine Lavoisier, a French scientist, politician, and public servant, during the period 1770 to 1794 made several contributions that paved the way for our modern view of atoms. He proposed a good theory for combustion, he showed that diamond is carbon and could be burned, and most significant of all, through his keen analytical sense he offered sound explanations and theories for what he observed. Lavoisier's *Traite*, which was comparable to Newton's *Principia*, in 1789 summed up much of his work and contains several concepts of greatest importance. For the first time, the principle of conservation of mass was effectively stated and explained. Although this principle, or law, which states that *matter can neither be created nor destroyed*, had been conceived earlier in philosophical thought, it was now derived experimentally. Lavoisier found that from 18 grams of grape sugar he got 8.8 grams of carbon dioxide and 9.2 grams of alcohol. In other words, he showed experimentally that the weight of the raw materials in a chemical reaction equals the combined weight of the products of the reaction, and in writing that

$$\text{grape sugar} = \text{carbon dioxide} + \text{alcohol}$$

he formulated an expression that is the forerunner of the

modern chemical equation. Countless other examples were found that, without exception, verified this law, which has become known as the *law of conservation of mass.*

It soon became obvious that mathematics could be applied to chemistry as it earlier had been applied successfully to physics. These attempts to find definite numerical values in chemical processes culminated in 1792 in a report by Jeremias Richter, a German chemist, in which he proposed that if a substance *AB* reacts with another substance, *CD*, both of whose compositions are known, then the composition of the products, *AD* and *CB*, could be calculated. Since his work was expressed in a rather difficult and involved style, it did not appeal immediately to other chemists. However, a few years later (1799) the French chemist, Joseph Proust, experimentally verified Richter's idea by showing that no matter how copper carbonate is prepared it always has the same chemical composition. Proust, after spending an additional nine years checking his conclusions, announced another important law—the *law of definite proportions*—which states that *when two or more elements combine, they do so in a fixed or definite proportion by weight.* For instance, when hydrogen and oxygen combine to form water, they always do so in the weight ratio of one part hydrogen to eight parts oxygen. When copper and oxygen unite to form black copper oxide, four parts of copper always combine with one part of oxygen. Figure 11-1 shows the proportions of zinc and sulfur that combine to form zinc sulfide—the ratio is slightly more than two parts zinc to one of sulfur. Note that if there is an excess of zinc above this proportional amount, it will be left over as free, unreacted zinc.

These two laws—the law of conservation of mass and the law of definite proportions—comprised the most important progress of the eighteenth century toward an understanding of the atomic concept of matter, and, along with the experimental evidence upon which they were based, laid the foundation for the next big step forward in chemical science —Dalton's atomic theory.

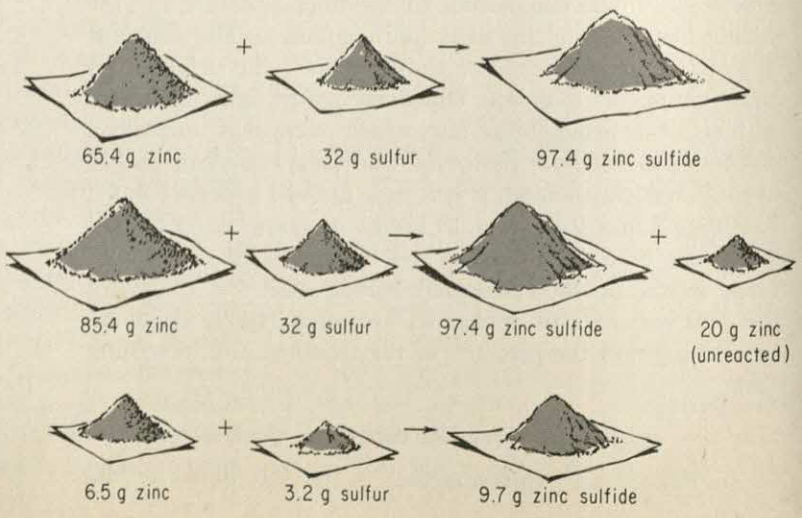


Fig. 11-1. An illustration of the law of definite proportions. Various quantities of zinc and sulfur may be added to each other, but the quantities that react will always be in a certain proportion.

11-3 Dalton's Atomic Theory

In 1808, John Dalton, an English schoolteacher, presented an explanation for these laws and the data behind them. His so-called atomic theory was an attempt to make a reasonable interpretation of the scientific data known at the time. Although his proposal was somewhat conjectural, it fit the facts and was gradually accepted as a valid theory. Dalton's conclusions were expressed in rather complex and ambiguous statements, but they boil down roughly to the following:

1. Atoms are distinct particles of matter that can neither be subdivided nor changed into atoms of a different substance.

2. Atoms are indestructible.

3. All atoms of a pure substance, or element, are identical.

4. Atoms of different substances have different weights.

5. When elements combine to form new substances—for instance, when hydrogen and oxygen combine to form water—the result is produced by the atoms of these elements reacting with one another in small, integral ratios, such as 1:2, 2:3, and so on.

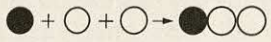
6. The relative weights of elements that combine with each other, according to the law of definite proportions, are proportional to the relative weights of the atoms of the elements. Thus, if five grams of calcium and two grams of oxygen react to form seven grams of calcium oxide, each calcium atom weighs two and a half times as much as an oxygen atom. The relative weights of these two elements will always be the same in this reaction regardless of how many atoms of each we use, simply because all calcium atoms are two and one-half times heavier than oxygen atoms.

The concepts summarized in Dalton's theory are the basis of our modern atomic theory. Their simplicity and the precision with which they explained the two laws above caught the interest and attention of his contemporaries, and, as a result, other scientists were stimulated to find the actual weights of the atoms themselves and to find more accurate quantitative relationships in chemical reactions. These activities culminated in several important advances, some of which we will now consider.

Dalton's theory led to the development of the *law of multiple proportions*, which is an extension of the law of definite proportions for those substances that contain the same elements in more than one ratio. For instance, carbon and oxygen combine in two ratios, 3:8 in carbon dioxide and 3:4 in carbon monoxide. The ratio of the oxygen in the first of these to the oxygen in the second is 2:1. That is, there are exactly twice as many oxygen atoms in carbon dioxide as there are in carbon monoxide. The law of multiple proportions may be stated: *Whenever certain elements combine in more than one ratio, there will be a simple, integral ratio*

among the amounts of one of these elements that is combined with a given amount of the other substances concerned.

In order to visualize the atoms that take part in chemical processes, Dalton invented a group of symbols, some of which are shown in Fig. 11-2. Thus, when he wished to show that one atom of carbon and two atoms of oxygen unite to form "carbonic acid" (now called carbon dioxide), he wrote:



This was much simpler than the old method of spelling out the names of the substances, and these symbols also more adequately relayed the idea that atoms of these substances are actually involved in the reaction. This type of notation was improved in 1814 by Berzelius, a Swedish scientist, who proposed the use of letters as symbols for the elements. The above reaction would then read:

$$C + O + O \rightarrow CO_2$$

Not only are the letters easier to write than Dalton's symbols, but they immediately suggest the element they represent because all elements are designated by one or two letters derived from either its present or its historical name. Hydrogen, for example, is H, oxygen is O, aluminum is Al, and magnesium is Mg. (See the atomic weight table on p. 216.) The subscripts are used to denote the number of atoms involved, there being only one atom when no subscript is used. These letter symbols have now become universally accepted and, like our system of numbers and measurement, are an international language in science.

In Dalton's time, approximately thirty elements were known. Since then, many more have been discovered, so that now our list numbers a little over one hundred. All the substances that occur in nature and that are produced in the laboratory are composed of these simple elements, and in view of the fact that over a million different substances have been identified, these relatively few elements must combine in many, many different ways to produce them all. Although the number of combinations that exist is very large, each of them has been found to be consistent with Dalton's theory. In other words, we have considerable evidence of the existence of atoms.

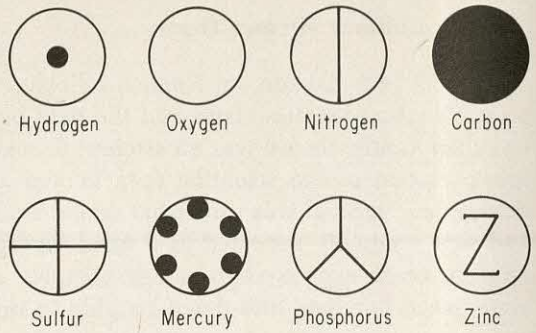


Fig. 11-2. Dalton's symbols for some of the elements.

11-4 Atoms and Molecules

When two or more elements combine to form a new substance, we call this substance a *compound.* Water, salt, and sugar are examples of compounds. A compound is represented by a collection of symbols—called a *formula*—each of which has a subscript that gives the exact number of atoms (if more than one) of that element in a unit particle of the compound. The formula for water, H_2O, represents the fol-

lowing information: (1) Water contains hydrogen and oxygen, and (2) a unit particle of water contains two atoms of hydrogen and one atom of oxygen. Similar information can be gleaned from the formula of any compound, such as NaCl for ordinary salt and $C_{12}H_{22}O_{11}$ for sugar. It is important to note here that H_2O does not mean that there are two parts by weight of hydrogen for each part by weight of oxygen. As Dalton pointed out, atoms of different elements have different weights, and from the formula alone we cannot say what the weight ratio between hydrogen and oxygen is unless we know the relative weights of the atoms of these two elements. We shall refer to these relative weights in a moment.

There was considerable difficulty, however, in determining the composition of the unit particle of a substance. In 1811, Amedeo Avogadro, an Italian, recommended that a new term be used to describe the arrangement and behavior of atoms in various substances, and, following his recommendation, the word *molecule* was introduced to represent the smallest particle of an element or compound. Avogadro suggested that ordinary oxygen is composed of molecules, each one containing two atoms, and that the gases hydrogen, nitrogen, and chlorine are also composed of diatomic molecules. Other elements such as sulfur may have even more complex molecules (S_8), while the metals iron and copper, and the inert gases—helium, neon, and so on—are composed of molecules that contain but one atom. A molecule is also the smallest particle of a compound substance. Thus, we speak of a molecule of carbon dioxide, CO_2, or a molecule of water, H_2O.

If Avogadro's concept had been accepted at the time it was proposed, it would have saved chemists 50 years of additional confusion. The history of this period from 1811 to 1860 is full of many attempts to reach a more adequate picture of the atom and molecule. There were bitter arguments, and no one could devise a completely satisfactory theory to explain the rapidly accumulating facts about atoms, their weights and chemical behavior. For instance, no one, including Avogadro, could offer an indisputable formula for water, methane (marsh gas), or ethyl alcohol. In fact, many chemists believed that water was HO until 1860. We will not dwell upon the details of this period, but instead will now go directly to the modern picture of the atomic composition of matter. In a later section we shall return to the means by which the formulas of many common substances were found.

11-5 Modern Atomic Theory

Although Dalton's atomic theory was generally accepted at the time, it has since been slightly modified to fit newly discovered facts. Atoms have been found to be composed of several parts, and, among the atoms of a particular element,

small variations in weights have been found. So conclusively, however, has the concept of atoms been proved that scientists today believe in the existence of them with as much confidence as they believe that the earth is round. Atoms are unbelievably small, light, and numerous. The face of a penny requires 10,000,000,000,000,000 atoms of copper to cover it with a single layer one atom thick, yet this enormous number of atoms weighs less than two ten-millionths of an ounce. No one has ever seen a single atom, but there is ample evidence that atoms have a measurable size and weight.

A single atom weighs in the neighborhood of 10^{-26} ounce. You may wonder how the weight of such a small object could be determined, but these incredibly small particles have been weighed and studied by many different means, and, surprisingly enough, there are at least fourteen different methods by which they have been accurately counted. One of these methods, briefly, is based on the fact that each atom possesses a certain number of unique electrons that are involved in many ordinary processes. For instance, when silver is plated onto an object, such as a teapot, each silver atom, in order to plate out of solution onto the teapot, must gain exactly one electron. Since we have a precise means of measuring the quantity of electrons in an electrical current, we can directly count the silver atoms. In a typical experiment, if we wish to plate out 107.9 grams of silver, 96,490 coulombs of electricity will be needed. Since each coulomb, as we know, contains 6.24×10^{18} electrons, the 107.9 grams of silver has 96,490 coulombs $\times$ 6.24×10^{18} atoms/coulomb, or 6.02×10^{23} atoms of silver. This particular quantity of silver, a gram atomic weight, is a *mole,* and the mole is defined as a quantity of matter containing 6.02×10^{23} particles of the substance concerned. Once we know the number of atoms in 107.9 grams of silver, it is a simple matter to calculate the weight of each atom.

The combined efforts of many scientists, working over many years, have finally produced an atomic theory that has been proved beyond a reasonable doubt to be valid, although we must keep in mind that it merely represents the best that is known now. Sometime in the future a newer and better theory may well be proposed, much as new theories have arisen in the past to supplant the older ones. Our present view of the atom can be outlined in this way:

1. All matter is composed of very small particles, or atoms.
2. Each element is composed of a different type of atom.
3. All the atoms of a particular element have the same or nearly the same weight.
4. Atoms can combine with almost all other types of atoms, in either simple or complex ratios, to form compounds.

It is noteworthy that Dalton's theory is very close to the above theory, although atoms are no longer viewed as inde-

structible and the atoms of a particular element are not exactly alike. And, with the study of more complex substances, the ratios by which atoms combine are not always small and simple. For instance, oxyhemoglobin, the red compound in blood, probably has the formula: $C_{738}H_{1166}O_{208}N_{203}S_2Fe$. Even many simpler substances such as salt, water, and sand, are more complex than the Dalton theory would indicate.

What these atoms actually are, how they react with one another, and what they are composed of are questions that we will consider in the next three chapters. Let us now examine a modern device that not only proves the existence of atoms, but also enables us to determine their weights.

11-6 The Mass Spectrograph

In 1919, in England, J. J. Thomson and F. W. Aston developed and perfected an instrument called a *mass spectrograph*. The principle upon which it operates may best be described by referring to an analogy. If you were to drop a mixture of golf balls and Ping-pong balls into the wind, the two kinds of balls would separate because they have different masses. That is, the Ping-pong balls would be deflected by the wind to a greater extent than the golf balls. If you knew the force of the wind, the force of gravity, and the distance through which the balls fell, you could determine the masses of the balls without weighing them directly. In a similar fashion, the atoms of an element can be projected from a chamber in a mass spectrograph by electrical means (Fig. 11-3), and then deflected by a magnetic and electrical field, which acts on the atoms as the wind did on the balls. The lighter atoms will be deflected more than the heavier ones, and, by accounting for the various factors that influence the paths of the atoms, scientists can calculate their mass without weighing them directly.

In 1912, Thomson experimented with neon in an early model of the spectrograph and, to his surprise, discovered that there are two different kinds of neon atoms and that one weighs about 10 per cent more than the other (Fig. 11-3). This fact is obviously at variance with Dalton's theory, which said that all atoms of neon should have the same

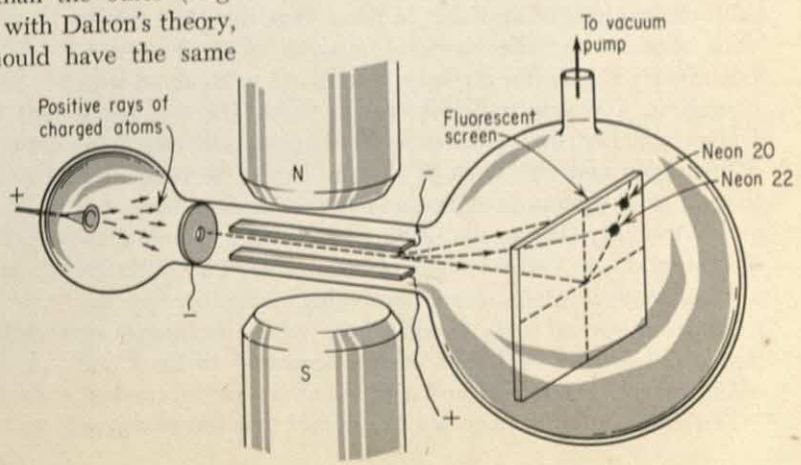


Fig. 11-3. The principle of the mass spectrograph.

weight. When an element has two or more kinds of atoms, we call these atoms *isotopes*. Other elements were analyzed with the mass spectrograph, and almost all of them were found to exist as mixtures of isotopes. Chlorine, for example, has two isotopes, iron has four, tin has ten, and mercury has seven. Even with the heavier atoms, such as uranium, the mass spectrograph is sensitive enough to differentiate between isotopes but differ in weight by less than 0.1 per cent. This instrument is now in wide use in the research laboratories of universities and industries throughout the world, and very large ones were used in connection with atomic energy production, in which the isotopes of uranium and other elements must be separated.

From the mass spectrograph, an oxygen atom was found to have a mass of 26.6×10^{-24} g, a hydrogen atom a mass of 1.66×10^{-24} g, and a uranium atom a mass of 395×10^{-24} g. Since it is inconvenient to deal with these very small weights, a more manageable system of weights, based on relative values, was devised even before the invention of the mass spectrograph. This system was developed through a study of the proportions by which atoms combine with one another, and, fortunately, the results of this study have been confirmed by the weight measurements of the mass spectrograph. Let us now turn back to two historical events that provided the means of finding the weights of atoms by indirect chemical methods. We have already mentioned one of them briefly—Avogadro's concept—and the other concerns a principle by which gases react with each other. We shall see how, even without the mass spectrograph, scientists were able to find the weight of single atoms.

11-7 The Law of Combining Volumes

As we have already noted, Dalton proposed that atoms of different elements have different weights. Since this idea explained the law of definite proportions, many of Dalton's contemporaries were spurred on to find the actual weights of atoms. Because it was impossible to weigh single atoms with the equipment available to these men, they had to use data obtained by the chemical analysis of various compounds. Hydrogen, for instance, was found to combine with oxygen in a weight ratio of one to eight. On this basis, Dalton assumed that a water molecule contained one atom of hydrogen and one atom of oxygen; hence, he concluded that an oxygen atom is eight times heavier than a hydrogen atom. Actually, Dalton was dealing here with a "combining" weight, and not a true weight of each atom. The difficulty became evident when these "combining" weights were derived from several compounds—there were inconsistencies that could not be explained. Atoms appeared to have one weight in one compound and a different weight in another.

The crux of the problem lay in the fact that there was no

way for Dalton and his contemporaries to know how many atoms of an element were in a compound. How could he determine that water has two atoms of hydrogen in each molecule, or that carbon dioxide has two oxygen atoms in it? Because this question could not be answered, it was merely assumed that one atom of one element reacted with one atom of the other, and since this assumption was false, obviously the conclusions were erroneous. The key to this problem was provided by the discovery of an important, yet simple, principle concerning the behavior of gases when they react chemically with each other.

In 1809, Joseph Gay-Lussac, a French chemist, observed that under identical conditions of temperature and pressure every time hydrogen and oxygen combine there is a constant relationship between the *volumes* of the two gases that are involved in the reaction. After repeated experiments, he found that two volumes of hydrogen always unite with one volume of oxygen, with an error of less than 0.1 per cent. When he inspected other gaseous reactions, he found that all gases combine in very simple volume ratios, such as 2 : 1, 1 : 1, 2 : 3, and so on. If we summarize the results of his experiments, we find that:

1 volume N gas + 1 volume O gas → 2 volumes nitric oxide gas
2 volumes H gas + 1 volume O gas → 2 volumes water vapor
3 volumes H gas + 1 volume N gas → 2 volumes ammonia
2 volumes CO gas + 1 volume O gas → 2 volumes carbon dioxide

These results are examples of what is now known as *Gay-Lussac's law of combining volumes: The volumes of gases entering into chemical changes are related to each other in simple, integral ratios.* In order to obtain these integral ratios, Gay-Lussac had to round off some of his figures slightly, but he firmly believed that an important relationship based on simple ratios existed here, even though his experimental data failed to prove it exactly. Here again, as with Galileo and his experiments on gravity, we see the intuitive ability of the scientist to overlook minor errors and difficulties in order to seek the basic principle that underlies the whole process.

Dalton refused to accept these results of Gay-Lussac, probably because they would destroy the simplicity of his theory. But no theory will last long unless it adequately explains the facts, and a new idea was shortly supplied to explain Gay-Lussac's combining volumes. Avogadro, as we have already noted, in 1811 suggested that gases may be composed of molecules rather than atoms, and that these molecules may contain one, two, three, or more atoms, depending upon the nature of the gas. His most important proposal, however—which came to be known as *Avogadro's law*—was that *equal volumes of gases at the same pressure and temperature contain the same number of molecules.* In other words, a liter of oxygen contains the same number of molecules as does a liter of hydrogen.

At the time, Avogadro's idea was no more than a happy guess. But let us see how it explains Gay-Lussac's law of

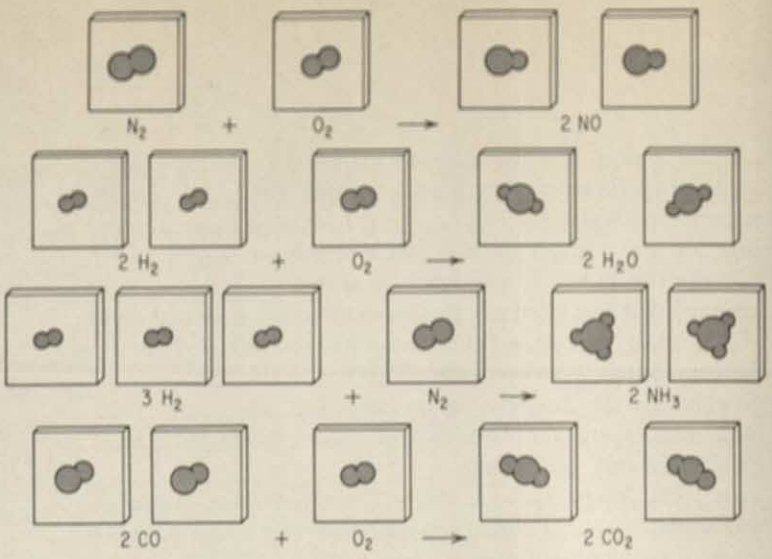


Fig. 11-4. An illustration of Gay-Lussac's law of combining volumes, showing the quantities of gases that react chemically with each other.

combining volumes, and at the same time would have supplied an answer to Dalton's problem of finding the true weights of atoms. Suppose we have a volume of gas just large enough to contain one molecule of gas. Although this is an extremely small volume, it will enable us to visualize the relationships among Dalton's atoms, Gay-Lussac's combining volumes, and Avogadro's molecules. In Fig. 11-4 each little box contains only one molecule of gas—the reactions illustrated are the same ones reported by Gay-Lussac. If we examine those reactions that involve oxygen, we note that one volume of oxygen, reacting with one volume of nitrogen, produces two volumes of nitric oxide. Since each molecule of nitric oxide contains some oxygen, where does all this oxygen come from? Avogadro proposed that the oxygen in the two volumes of nitric oxide, coming from only one volume of oxygen, must have come from oxygen molecules that contained two atoms of oxygen. If this is true, then the hydrogen gas molecule probably has two atoms (diatomic), and since two volumes of hydrogen make only two volumes of water vapor, each molecule of water must contain two atoms of hydrogen. From this we can readily deduce the relative weights of the single hydrogen and oxygen atoms, that is, 1:16 instead of 1:8.

So radical were Avogadro's ideas and so lacking in experimental proof that his suggestions were received coolly by his contemporaries. In fact, it was only after many years of additional experimental and theoretical developments that Avogadro's ideas were revived. Another Italian scientist, Stanislao Cannizzaro, had been teaching Avogadro's hypothesis to his students at Genoa for some time, and when the leading chemists were called together at the first International Chemical Congress in 1860, he suggested at the close of this meeting that others might profit through a review of Avogadro's ideas. He pointed out that the weight of an oxygen atom is only half the weight of its molecule, and so on. Thus, after fifty years, Avogadro's work became recognized as accurate and dependable, and it was only a short time before the confusion about the weights of atoms and the

writing of formulas and equations was cleared up. In the next section we shall consider the weights of atoms, and in Table 11-1 we shall see the difference between the weights of molecules and the weights of atoms.

11-8 Atomic Weights

We now have arrived at a point where a dependable set of weights for the atom can be derived. Instead of making a list of the actual weights of the atoms, it has been more convenient to collect a set of relative weights that leads to a system of atomic weights for the elements—a system that has become internationally used and that is periodically reviewed by the International Union of Pure and Applied Chemistry. The *atomic weight* of an element is the weight of an atom of that element compared to the weight of an atom of the more abundant isotope of carbon, C-12. Since 1961, the arbitrary reference point in the system of atomic weights has been the C-12 isotope to which has been assigned 12 atomic weight units. It was selected as the standard reference primarily for convenience and to eliminate the confusion that once existed when there were two systems, one for physical purposes and one for chemical purposes. Table 11-1 summarizes the atomic weights of some of the naturally occurring elements and the molecular weights of a few compounds.

Ordinary carbon, as it occurs naturally, is composed of two isotopes, C-12 and C-13. This isotopic mixture has an atomic weight of 12.01115, which indicates that only a very small amount of C-13 exists in nature. Atomic weights as based on chemical analyses provide no clue to the presence of isotopes—only the measurement of atomic weights by physical methods such as the mass spectrograph reveal the existence of isotopes.

11-9 Chemical Equations

In physical science it is often necessary to use symbols to communicate ideas. We have already had occasion to express concepts in terms of mathematical equations. Now we shall find that all known reactions of the elements and their compounds can be expressed in a shorthand fashion called *chemical equations*. Although at first you might think that these equations are unintelligible and written to complicate matters, they are actually used as a matter of convenience and, indeed, are necessary. You will recall that one of the important features of the scientific method is the publication of results. These reports must be understandable and concise, and equations provide an effective, universal language that makes it possible for others to extend their scientific knowledge. A chemical equation is a collection of symbols and

TABLE 11-1 ATOMIC WEIGHTS OF NATURALLY OCCURRING ELEMENTS BASED ON THE ISOTOPE CARBON-12 = 12.0000

Element	*Symbol*	*Atomic weight*	*Molecule*	*Relative weight of molecule*
Hydrogen	H	1.00797	H_2	2.0159
Helium	He	4.0026	He	4.0026
Carbon	C	12.01115	C	12.0112
Nitrogen	N	14.0067	N_2	28.0134
Oxygen	O	15.9994	O_2	31.9988
Fluorine	F	18.9984	F_2	37.9978
Sodium	Na	22.9898	Na	22.9898
Aluminum	Al	26.9815	Al	26.9815
Silicon	Si	28.086	Si	28.066
Sulfur	S	32.064	S_8	256.512
Chlorine	Cl	35.453	Cl_2	70.906
Calcium	Ca	40.08	Ca	40.08

Compounds	*Formula*	*Relative weight of molecule*
Water	H_2O	18.0153
Carbon dioxide	CO_2	44.0091
Ammonia	NH_3	17.0234
Nitric oxide	NO	30.0061
Sulfur dioxide	SO_2	64.0628

formulas that describes a chemical reaction. The substances that enter into the reaction, called *reactants*, are written on the left-hand side, and the *products* of the reaction are written on the right-hand side. An arrow is usually used to indicate the course of the reaction. In Fig. 11-4 beneath each group of boxes, chemical equations, derived from experimental evidence, have been written describing each reaction.

Let us take another experimental observation and express it in a chemical equation. When heated, mercury oxide, a red powder, decomposes into mercury metal and oxygen gas. This reaction can be written:

$$\text{mercury oxide} \rightarrow \text{mercury} + \text{oxygen}$$

Substituting formulas for these terms, we have:

$$HgO \rightarrow Hg + O_2$$

An examination of this expression reveals that there are two atoms of oxygen on the right and only one on the left. According to our atomic theory and to the law of conservation of mass, no atoms should be lost or created in a chemical reaction. Since this reaction was proved experimentally, we must do something to the equation to make it fit the facts. We must, in short, balance it by supplying suitable coefficients, or numbers, before each of the terms so that an equal number of atoms of mercury and oxygen will appear on each side of the equation. The simplest arrangement is always preferred, and, therefore, the balanced equation will be:

$$2HgO \rightarrow 2Hg + O_2$$

This is now a balanced equation, and it tells us that two molecules of mercury oxide decompose to form two molecules of mercury and one molecule of oxygen, which you will recall, is a diatomic molecule.

To take another example, the reaction between zinc chloride and hydrogen sulfide produces zinc sulfide and hydrogen chloride:

$$\underset{\text{zinc chloride}}{ZnCl_2} + \underset{\text{hydrogen sulfide}}{H_2S} \rightarrow \underset{\text{zinc sulfide}}{ZnS} + \underset{\text{hydrogen chloride}}{HCl}$$

This equation is unbalanced, because unequal numbers of hydrogen and chlorine atoms appear before and after the reaction, so we must supply the following coefficient:

$$ZnCl_2 + H_2S \rightarrow ZnS + 2HCl$$

Some practice and experience is required to attain proficiency in writing chemical equations, but the above explanation should be sufficient for our present purposes. It is advisable, however, to become acquainted with chemical notations to the extent that you can understand what a formula and an equation mean. For example, from the equation:

$$SiO_2 + 4HF \rightarrow SiF_4 + 2H_2O$$

you should be able to read the following information: a molecule of silicon dioxide reacts with four molecules of hydrogen fluoride to produce a molecule of silicon combined with fluorine (silicon tetrafluoride) and two molecules of water. Furthermore, this equation tells us that a molecule of silicon dioxide contains one silicon atom and two oxygen atoms, a molecule of hydrogen fluoride contains one hydrogen atom and one fluorine atom, and a molecule of silicon tetrafluoride contains one silicon atom and four fluorine atoms.

The naming of compounds follows a few simple rules. The elements are usually named in the order they appear in the formula. If the molecule contains only two different elements, the ending of the second element is replaced with the suffix *-ide*. If there are three elements in a compound, one of them is usually oxygen. If the compound is a common one, the other two elements are named and the ending of the second element is replaced with the suffix *-ate*. Here are some examples of these naming rules:

$NaCl$	sodium chloride
CO_2	carbon dioxide
CO	carbon monoxide
CaS	calcium sulfide
$NaNO_3$	sodium nitrate
$NaClO_3$	sodium chlorate
$CaSO_4$	calcium sulfate
$AlPO_4$	aluminum phosphate

A chemical equation also tells us the quantities of each substance that will react and the amount of each product

that is formed. Let us examine the equation:

$$Ca + S \rightarrow CaS$$

Since we know the relative weights of the calcium and sulfur atoms (Ca = 40 and S = 32) we can translate those atomic weights into any convenient weight units that we wish. This equation tells us that one calcium atom reacts with one sulfur atom. Hence, 40 grams of calcium would react with 32 grams of sulfur because there would be the same number of calcium atoms in 40 grams of calcium as there are sulfur atoms in 32 grams of sulfur. In this case we would get 72 grams of calcium sulfide. If we wish, we may express these quantities in other weight units—kilograms, pounds, or tons, as long as we use the same units throughout any particular situation.

11-10 Atoms and X-ray Diffraction

In Chapter 10 we saw how X rays, extremely short electromagnetic waves, were produced by electrons striking a metal target in a cathode ray tube. Within 25 years after X rays were discovered, it was observed that they could be reflected and refracted like ordinary light waves, but with one notable exception, and this exceptional property turned out to be one of the most useful applications to which X rays have been put.

The wave length of X rays is about one or two Angstrom units—a distance very nearly the size of atoms themselves. Max von Laue of Germany predicted that when an electromagnetic wave is reflected from a surface made up of atoms whose size is nearly the same as the wave length of the wave, an exceptional effect, called *interference*, will be observed. Interference occurs when water waves strike objects whose size is about the same as the wave length of the waves. Or, we can see interference of water waves as they pass through adjacent openings in a jetty. Figure 11-5 shows water waves passing through three openings in such a jetty. The waves are reflected from the solid jetty, but pass through the three openings and then fan out in a circular fashion. There are now three sets of waves, each having the same wave length but starting from three different positions. When these three waves meet, they interfere with each other—when the crest of one wave and the trough of the other arrive at a point at the same time, they cancel out each other. But when the crests of the waves arrive at a certain point at the same time, they reinforce each other, and the same holds for two troughs that meet at a certain point.

From Fig. 11-5 we note that reinforcement of the waves occurs at certain angles to the jetty. These angles are directly determined by two factors—the distance between the openings in the jetty and the wave length. This relationship, which will shortly be discussed in detail, is known as *Bragg's law of diffraction*, after Sir William Bragg and his son, Sir

Fig. 11-5. Water waves, passing through openings in a jetty, undergo interference.

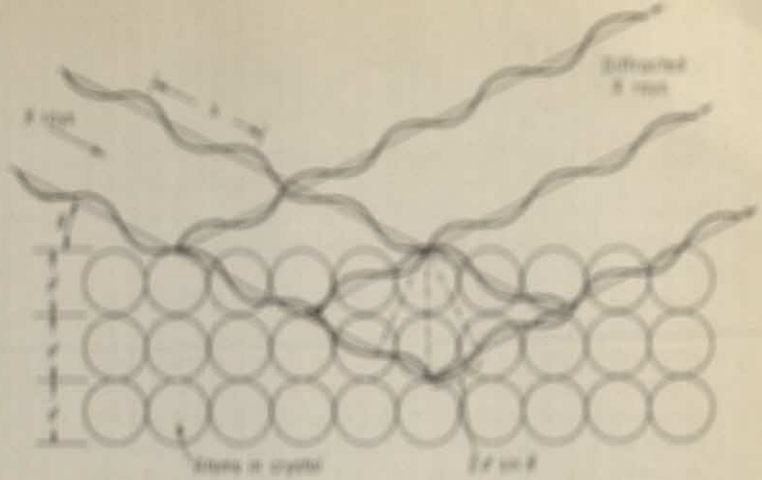


Fig. 11-6. The diffraction of X rays by the atoms in a crystal, showing the relation between the wave length of the X rays, the angle of reflection, and the distance between the layers of atoms in the crystal.

Lawrence, of England, who developed it following the original suggestion of von Laue. Let us now return to the effect of atoms upon X rays.

When X rays strike a crystal, they undergo interference because the atoms are about the same size as the wave lengths of X rays. In Fig. 11-6, we see X rays approaching the surface of a crystalline solid, wherein the atoms are arranged in a regular, geometric pattern. As the X rays are reflected from the atoms, they scatter out in a circular fashion as though the X rays had started from the atoms themselves. Just as the water waves interfere at some angles and reinforce at others, the X rays will interfere with each other at certain angles and reinforce each other at other angles. In the figure, the angle at which the X rays approach the crystal is adjusted so that the X rays that are reflected will be reinforced at the same angle as the ordinary angle of reflection. This angle can be precisely measured because the instrument that is designed for this study—the X-ray diffractometer—records the angles at which X rays are reinforced as shown in Fig. 11-7. Mathematical analysis of the geometry of the crystal lattice in Fig. 11-6 gives us the Bragg equation:

$$\lambda = 2d \sin \theta$$

In this equation λ is the wave length of the X ray, d is the distance between the atoms or molecules in the crystal that cause interference, and $\sin \theta$ is a quantity related to the

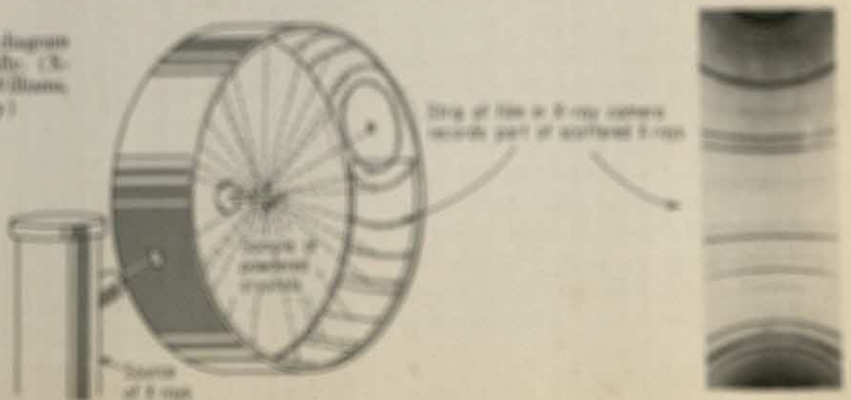


Fig. 11-7. An X-ray diffraction diagram recorded photographically. (X-ray diagram by M. B. Williams, Oregon State University.)

angle of diffraction. The derivation of this equation need not concern us here, but from it we can readily see that the wave length of the X ray is directly proportional to the distance between the particles in the crystal and the angle at which reinforcement, or diffraction, occurs.

In order to use X rays to study crystals we must know their wave length, which can be found if we have a crystal whose structure is known. Fortunately, there are several single crystal systems whose structures are known and from which the wave length of an X ray can be measured. For instance, the X rays emanating from a copper electrode in a cathode-ray tube have a wave length of 1.5374 A. Once the wave length is known, X rays can be used to study the distances between atoms and molecules in the most complex solids. Let us see how such an experiment is conducted.

If we place a single crystal of sodium chloride, or common salt, in a beam of X rays from a copper X-ray source and rotate the crystal carefully in order to find the angle at which there is maximum reinforcement of the diffracted waves, we get an angle of 15 degrees and 50 minutes, almost the same as the angle indicated in Fig. 11-6. Since we know the wave length of the X rays and the angle of diffraction, we can now calculate the distance d between the atoms in the salt crystal:

$$\lambda = 2d \sin \Theta$$
$$1.5374 \text{ A} = 2d \sin 15^\circ\ 50 \text{ min}$$

From mathematical tables, we find that sin 15° 50 min is 0.273, hence:

$$2d = \frac{1.5374}{0.273} = 5.64 \text{ A}$$

Therefore d, the distance between the sodium and chlorine in the crystal, is 2.82 A.

We can now use this information to count the atoms in a salt crystal. For instance, let us assume that we have a salt crystal weighing 58.5 grams. This particular weight of sodium chloride is the sum of the atomic weights of sodium (23.0) and chlorine (35.5) expressed in grams. This is called a *gram molecular weight* and is a mole. Whenever we add up the atomic weights of the elements in a compound and express this number in grams, we get its gram molecular weight. For instance, a gram molecular weight of water is 18 g, of carbon dioxide (CO_2) 44 g, and of sugar ($C_{12}H_{22}O_{11}$) 342 g. This unit is an extremely useful tool for the chemist because it gives him a large enough quantity of the substance for experiments that are directly related to the atomic weights of the elements concerned.

One salt crystal weighing 58.5 grams makes a cube almost exactly 3 cm on an edge. If the distance between rows of particles in the crystal, as revealed by X-ray diffraction, is 2.82 A and if these particles are arranged in the cubic patterns shown in Fig. 11-6, we can calculate the number of

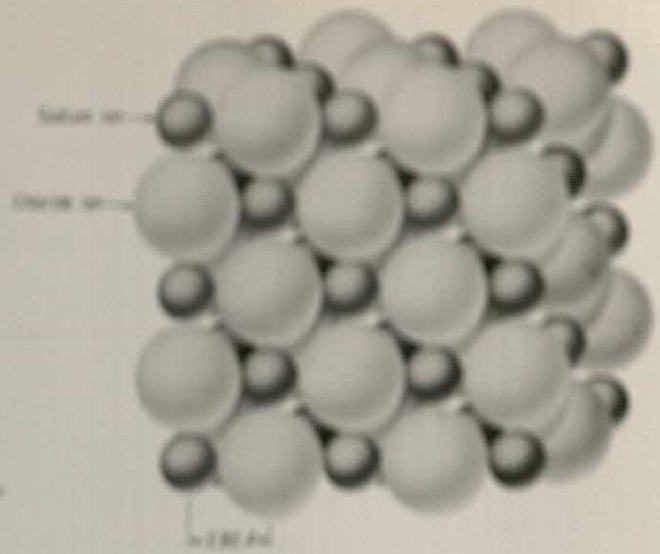


Fig. 11-8. Diagram of the sodium chloride crystal.

atoms on each edge of the cube:

$$\frac{3.00\text{ cm}}{2.82 \times 10^{-8}\text{ cm}} = 1.064 \times 10^{8}\text{ particles}$$

If there are 1.064×10^8 particles on each edge of the cube, in the whole cube itself there will be $(1.064 \times 10^8)^3$, or 1.204×10^{24} particles. Since half of these are sodium and half are chlorine, there are 6.02×10^{23} atoms of sodium and 6.02×10^{23} atoms of chlorine. This number, 6.02×10^{23}, is called *Avogadro's number* in honor of the man who first proposed that equal volumes of gases contain equal numbers of molecules. Although Avogadro was not able to count the actual number of atoms or molecules in a gram molecular weight, he is credited with the idea that there was a definite number. Since his time, scientists have counted atoms many times by several different methods (one of which is the electroplating method described earlier in this chapter), and in every instance they have found that a gram molecular weight of any substance contains 6.02×10^{23} molecules. Thus, we see that X rays have not only demonstrated the existence of atoms—how else could they be diffracted except by tiny particles about 1 or 2 Angstroms in size?—but have provided another method by which we can count atoms themselves.

11-11 Summary

Atoms, the unit particles that compose all matter, were first conceived in philosophic thought and have now been proved to exist by various methods. The philosophers of Greece and the early experimenters who brought about the chemical revolution believed that atoms were the primary units of matter, even though they could not experimentally verify their existence.

The first proofs of atoms were based on the law of definite proportions, the law of conservation of matter, Gay-Lussac's hypothesis, and Avogadro's hypothesis. Dalton's atomic

theory, with certain modifications that were required by experimental findings in later years, gives us a working explanation of the behavior of atoms.

Atomic weights are the relative weights of atoms based on an arbitrary reference standard of 12 atomic weight units for C-12. Molecular weights are the sum of the atomic weights of the elements in a substance.

Chemical equations efficiently express the course of a chemical reaction through a collection of symbols that record the reactants and the products in the reaction.

Some of the most conclusive evidence that atoms exist is exhibited by X-ray diffraction of crystalline solids. In a gram molecular weight of a substance there are Avogadro's number of molecules—6.02×10^{23}.

EXERCISES

A. KEY TERMS TO REMEMBER

atom
atomic weight
Avogadro's law
Avogadro's number
chemical equation
compound
element
gram-molecule
interference
isotope
law of conservation of mass
law of definite proportions
mass spectrograph
mole
molecular weight
molecule
reactants and products
X-ray diffraction

B. QUESTIONS ABOUT ATOMIC THEORY

1. State the modern interpretation of Dalton's theory and give several ideas that it contains.

2. What parts of Dalton's theory have been modified by later discoveries?

3. Explain the law of definite proportions in terms of Dalton's atomic theory.

4. Describe the operation of a mass spectrograph. How can this instrument be used to weigh an atom?

5. Why was there a delay of many years before Avogadro's principle was accepted?

6. What is the standard used to determine atomic weights? Why is this particular standard used?

7. If the isotopes of an element are always found mixed in the same proportions, what proof do we have that they exist?

8. By referring to Table 11-1 and the list of compounds in Section 11-9, name the following compounds: Na_2O, SiO_2, Al_2S_3, $Ca(NO_3)_2$, $CaCl_2$.

9. Give in words the meaning of the following equations:

$$2Ca + O_2 \rightarrow 2CaO$$

$$C + O_2 \rightarrow CO_2$$

$$2Al + 3Cl_2 \rightarrow 2AlCl_3$$

10. Describe the way X rays are used to study the arrangement of atoms and molecules in a solid.

11. Tell briefly how we may count the atoms in a substance if we know its density, the distance between its atoms, and the weight of a gram-molecule of it.

12. List all the experimental evidence you can that proves that atoms exist.

C. PROBLEMS ABOUT ATOMIC THEORY

1. Using the information in Table 11-1, calculate the molecular weights of sodium chloride (NaCl), sulfuric acid (H_2SO_4), and sugar ($C_{12}H_{22}O_{11}$). Ans. 58.5 for NaCl, 98 for H_2SO_4.

2. From the equation, Ca + S → Cas, calculate the weight of calcium needed to make 100 grams of calcium sulfide. Ans. 55.6 g.

3. How many molecules of water are in a glass of water (250 cm^3)? Ans. 83.2×10^{23}.

4. If the molecules in a glass of water were divided among the population of the world, how many molecules would each person have? Assume world population is three billion.

5. Natural rubidium is composed of two isotopes, Rb-85 and Rb-87, present to the extent of 72.3 per cent and 27.7 per cent, respectively. Assuming that the atomic weights of each isotope are whole number values as given, what is the atomic weight of the naturally occurring element?

6. Assuming that the diameter of a gold atom is 3.0 A, calculate the number of gold atoms needed to cover one square centimeter of surface with a layer one atom thick and with each atom touching the ones adjacent to it.

Twelve

Atomic Structure

In Chapter 8 we examined briefly the electrical nature of matter. We saw that the negative charge formed on a rubber rod when it is stroked with fur is caused by the excess *electrons* that are borrowed from the fur. The positive charge on the fur is inherently within the atoms and molecules of the fur, and we called the single positive charges *protons*. In addition to the electrons and protons, atoms contain neutral particles called *neutrons*. We will now investigate the manner in which these three kinds of particles make up atoms and determine their behavior.

The study of atomic structure is an intriguing and broad one. You will have occasion to draw upon your knowledge of electrostatics, electricity, magnetism, wave motion, light, X rays, and radioactivity. The basic concepts of space, time, force, and energy will be applied. Although you will need these tools and many more to explore fully the structure of these miniature worlds, once you gain a picture of the inside of the atom you will find that the behavior of all atoms is extremely consistent with this fundamental view.

Why is iron magnetic, why does salt taste salty, what makes a neon sign glow, and why is copper a good conductor of electricity? Why is diamond hard, what makes iron rust, and why in Siberia does tin turn to dust? The answers to these questions are based on the structure of the atom, and the more familar we become with the atom the less mysterious and more understandable the answers become. It has been said that if we knew how the parts of an atom of a substance function, we could predict the behavior of the substance at all times. We have a lot to learn about the atom, but we know enough now to explain much of the physical and chemical behavior of matter.

Great strides were made in unraveling the secrets of the atom in the ten-year period from 1895 to 1905. Roentgen announced the existence of X rays in 1895, Becquerel found radioactivity in 1896, the Curies found polonium and radium, and in 1905, Einstein confirmed the existence of the *photon*, a unit quantity of energy predicted by Max Planck in 1900. In 1902, Rutherford and Soddy began a series of experiments on radioactivity that literally unlocked the innermost parts of the atom. Seldom in the history of science have so many significant advances taken place in so few years—a period often called "the golden age of atomic science." To introduce this study, let us examine briefly some of these important events—events that provide the framework for the modern picture of the atom.

12-1 X Rays and Radioactivity

Becquerel made his famous "accidental" discovery while investigating the fluorescence that X rays produce in glass and various minerals. He already knew that X rays penetrate many common substances, and physicians in Vienna used these rays to assist in surgery only a few weeks after Roentgen's discovery. While trying to find a relationship between the fluorescence of minerals and their effect upon photographic film shielded from ordinary light, he by chance placed some of his film near a sample of uranium salt that had not been excited, either by X rays or by sunlight. To his surprise, he found that uranium emanated some sort of radiation that affected the photographic film in about the same way X rays did.

At this point, Marie Sklodowska Curie started a systematic study of "Becquerel rays" and identified several substances that showed even greater effects upon photographic film than did uranium. In 1898, Mrs. Curie, along with her husband Pierre Curie, announced first the discovery of polonium and then of radium. Both elements were discovered and separated from their ores by the detection of their radiation. This radiation was found to consist of three distinct types, which can be separated and identified by deflecting them through a magnetic or electrical field (see Fig. 8-9). By this means, the three types were found to be positively charged particles (alpha rays), negatively charged particles (beta rays), and a neutral radiation that is unaffected by the magnetic field (gamma rays).

Further study of these rays revealed that alpha particles bear two unit positive charges and have a mass four times that of the hydrogen atom. Beta particles are electrons that are emitted by radioactive elements, and gamma rays are electromagnetic rays similar to X rays. The properties of these particles, along with the electron, proton, and neutron, are summarized in Table 12-1. Numerous other particles have been discovered that explain certain features of atomic

TABLE 12-1 FUNDAMENTAL PARTICLES

	Charge	Rest Mass (electron = 1)	Rest Mass (atomic wt. scale)	Rest Mass
Electron	−1	1	1/1836	9.11×10^{-28} g
Proton	+1	1836.1	1.00755	1.673×10^{-24} g
Neutron	0	1838.6	1.00894	1.675×10^{-24} g
Alpha	+2	7302.2	4.0026	6.651×10^{-24} g
Beta	−1	1	1/1836	9.11×10^{-28} g

structure, and these will be considered in detail in Chapter 25 (Table 25-1). It is of interest to note here that *mesons* of various sorts (whose masses are about 250 times that of an electron), *neutrinos* (whose mass and charge are negligible), and *positrons* (positively charged electrons) have been indicated in experimental and theoretical studies. For our present purpose, however, the particles in Table 12-1 will be sufficient to introduce us to atomic structure.

12-2 Inside the Atom

In spite of the incredibly small size of atoms, the particles of which they are made are still smaller. Rutherford, intrigued with the problem of what the inside of the atom is like and how large some of the inner parts of the atom are, devised an ingenious experiment. He bombarded the gold atoms in a thin sheet of gold foil with alpha particles from a radioactive source, in a process that was like using a shotgun to hit a small target. His "gun" was a block of lead in which a small hole was bored (Fig. 12-1). At the back end of the hole he placed a small quantity of radium, which ejected a constant supply of alpha particles out of the open end of the hole. When the stream of alpha particles was directed against the thin sheet of gold metal, most of the particles passed right through the gold foil. In a darkened room, Rutherford observed the path of the alpha particles by placing a fluorescent screen behind the gold foil (Fig. 12-1), and each particle, as it struck the screen, produced a small flash of light. From a count of the particles that passed through the gold atoms, he found that only one alpha particle in 100,000 was deflected when the stream passed through 1000 atoms of gold; the great bulk of the particles passed through with no change in direction.

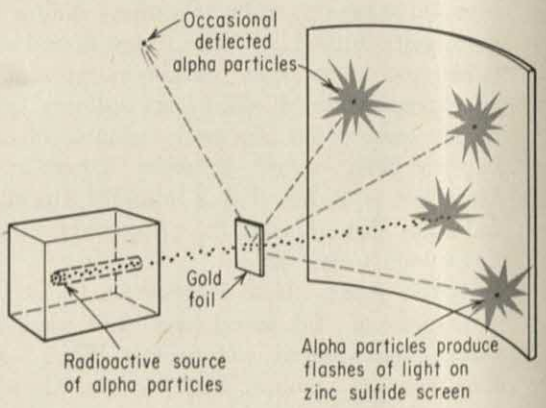


Fig. 12-1. Rutherford's experiment of bombarding a piece of gold foil with alpha particles.

Rutherford concluded from this experiment that the atom consists of two regions—a sparsely populated region that contains nothing capable of deflecting an alpha particle and that accounts for the greater part of the atom, and a very dense region of extremely small size. We call this compact portion of the atom the *nucleus*, and we deduce that it is positively charged because it is capable of deflecting alpha particles. The outer region of the atom contains only elec-

trons, whose mass is insufficient to cause a deflection of alpha particles (Fig. 12-2). Rutherford calculated that the diameter of the gold atom nucleus is only $\frac{1}{10,000}$ of the diameter of the whole atom. It is easy to see then why an alpha particle seldom passes near enough to the nucelus to cause its direction to be changed. The situation is much the same as if we were to shoot a BB gun at a mosquito flying around in a football stadium—it would require each of the 100,000 fans in the stadium firing 1000 BB's at the same time to score a probable hit.

The nucleus of an atom is an extremely dense quantity of matter. Since most of the mass of the atom is concentrated in the nucleus, which is approximately 10^{-12} cm in diameter, the density of a collection of atomic nuclei would be very high. For instance, a cubic inch of gold nuclei would weigh over 100 million tons, but a cubic inch of ordinary gold atoms weighs a little less than one pound. Matter has been found in places to be much more dense than the atoms we ordinarily encounter. The star, Sirius B, as well as many others, has a density of over one ton per cubic inch. This high density indicates that these stars may be composed of collapsed or compressed atoms.

Suppose an atom were magnified to the size of a basketball. Even then, we still would be unable to see its electrons, and the nucleus would be only about .001 cm in diameter, or one-tenth the thickness of a sheet of paper. Therefore, it is quite amazing to find that matter as we ordinarily know it—gold, iron, wood, glass, and even our bodies—is composed of atoms that are mostly space. Teachers who imply that our heads are empty may be nearly correct after all!

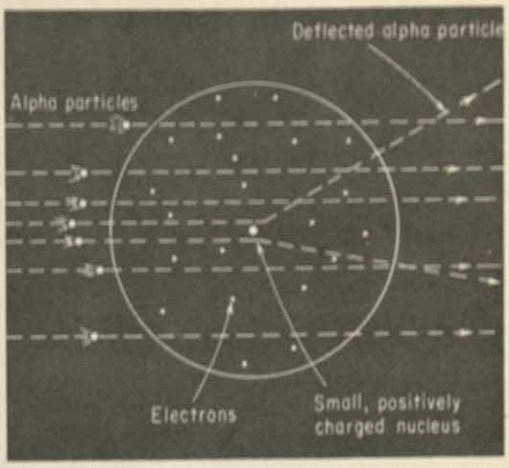


Fig. 12-2. The picture of an atom derived from the Rutherford experiment.

12-3 The Simplest Atom

In the last chapter we learned that hydrogen is the lightest element known, and from this fact we deduce that it is the simplest atom that exists. A hydrogen atom contains the least number of electrons, protons, and neutrons of all the atoms that we know; it contains one proton in its nucleus and one electron in its outer portion.

But how can the electron, with its negative charge, and the proton, with its positive charge, remain so near each other in a hydrogen atom and not fall together? There must be some force that holds the two tiny particles apart, a force equal and opposite in effect to the coulombic force of attraction between the two unlike charges. Atoms have been thought of as spherical particles for a long time, so it was only natural for the pioneers of atomic structure to conceive of the hydrogen atom as a miniature solar system, with the electron traveling in an orbit around the proton at the center of the atom. Niels Bohr, a Danish physicist, in 1913 refined this view and provided us with the first adequate working model of the structure of the atom.

Although Bohr's orbital concept of the atom has since been modified, it is accurate enough to begin our study of the atom. In the hydrogen atom, the electron travels with a speed of 2.18×10^8 cm/sec, or about 1 per cent of the speed of light. The electron does not exactly follow an orbit like that of a planet around the sun, instead it travels in a wave-like motion, moving toward, and then away from the nucleus. In effect, the electron sweeps over all the space from the nucleus out to about 1 A and thus gives the atom a spherical shape. The most likely or average place that we would find the electron in the hydrogen atom is at a distance of 0.53 A from the nucleus. (This is the apparent size of the hydrogen atom based upon a study of X-ray diffraction and the density of a large number of hydrogen atoms, such as that shown in the last chapter for sodium chloride.) A cross section of the hydrogen atom would therefore look something like Fig. 12-3a, with each dot representing the probability of finding the electron at any given instant. Because of the unpredictable position of the electron in the atom, a description of the behavior of the electron cannot be given in terms of ordinary geometry. Instead, rather advanced mathematics, including the concept of probability (which we will take up in Chapter 30), are applied to the electron's location, and for this reason our picture of the atom will be somewhat crude. As a matter of convenience, this picture of the atom

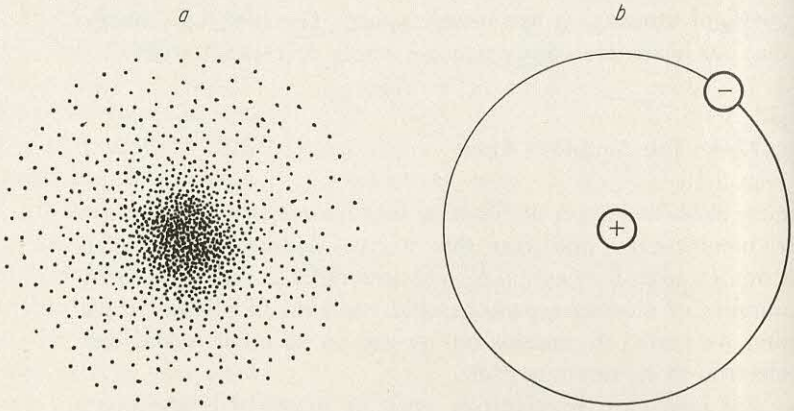


Fig. 12-3. Two pictures of the hydrogen atom.

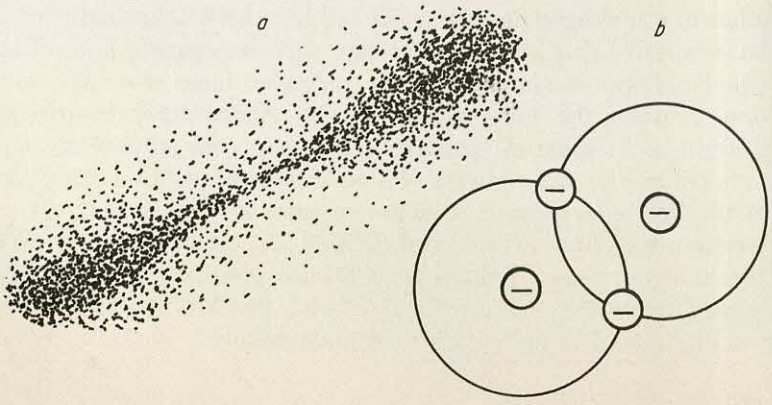


Fig. 12-4. Two pictures of the hydrogen molecule, after the manner shown in Fig. 12-3.

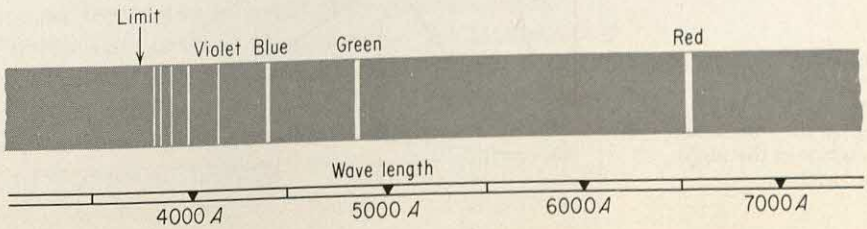
can be illustrated by the diagram in Fig. 12-3b, where the circle encloses the volume in which we might expect to find the electron under ordinary conditions.

When two hydrogen atoms are combined in a hydrogen molecule, H_2, the electrons become associated with both nuclei. Although the nuclei, being positively charged, cannot approach each other too closely, the orbital energy of the two electrons tends to overcome the coulombic repulsion of the two nuclei and to hold the positively charged particles about 0.75 A apart. The H_2 molecule can thus be described by the diagrams in Fig. 12-4. Here again, it is convenient to depict the hydrogen molecule as two spheres that overlap slightly. The mutual sharing of the two electrons by the two hydrogen nuclei is also indicated in the simpler diagram.

12-4 The Hydrogen Spectrum and the Photoelectric Effect

As early as 1880, physical scientists studied the light produced by various gases when they are excited by an electrical discharge through a Crookes tube (Fig. 8-5). A small amount of a gas, such as hydrogen, neon, nitrogen, etc., is admitted to the evacuated tube, and upon passing an electrical discharge through the tube the gas emits various colors of the spectrum that can be readily analyzed by passing the light through a prism spectroscope (Fig. 10-24). The light from glowing hydrogen will appear as a series of colors—red, green, blue, and shades of indigo and violet (Fig. 12-5). Neon emits an orange-red color of light familiar to us in the modern advertising sign, and upon spectroscopic examination, neon's light would be found to contain many specific spectral lines. Each element gives off its own set of colored lines when it is excited, and as a result, this phenomenon can be employed as a means of detecting even small amounts of an element. So dependable is the color of the light produced by the various elements that the wave length of the orange-red line in krypton's spectra has become the basis

Fig. 12-5. The hydrogen spectrum. The red line on the right corresponds to the transition of the electron from the third to the second energy level. The green line corresponds to the transition of the electron from the fourth to the second level. The limit of the lines in this portion of the spectrum corresponds to the transition of an electron from the highest possible energy level to the second energy level.



for defining the length of a meter. On this basis, the meter is 1,650,763.73 times the wave length of this particular spectral line of krypton.

In order to understand how the hydrogen atom, or krypton and other elements, emit certain colors of light unique to the particular element, we must digress a bit to consider the work of Max Planck and Albert Einstein during the years around 1900 to 1905. Upon investigating the energy associated with light of various colors, Planck noted that the radiant energy emitted by very hot bodies was of shorter wave lengths than that emitted by cooler bodies. The graph in Fig. 12-6 shows the range of wave lengths emitted by three bodies at various temperatures, and it is noted that the most probable radiant energy from the sun at 6000°K has a wave length of about 4800 A, in the green portion of the spectrum. A glowing lamp filament at 4000°K, on the other hand, emits radiant energy at a most probable wave length of about 7000 A, in the red portion of the spectrum. Based on this experimental relationship, Planck suggested that radiant energy is inversely related to its wave length, or directly related to its frequency in the simple equation:

$$E = hn \qquad (12.1)$$

E is the energy in joules of radiation whose frequency is n, and the proportionality constant h, known as Planck's constant, is 6.6256×10^{-34} joule sec.

Mechanistically, Planck's equation describes light as a stream of energy bundles, sometimes called *photons*. The energy of a photon of violet light is greater than the energy of a photon of red light because the bundles of energy in violet light are greater in frequency than those in red light. The energy bundles of X rays and cosmic rays possess sufficient energy, because of their extremely high frequency, to penetrate many materials that are opaque to visible light.

A more complete explanation of light as bundles of energy, known as Einstein's theory of the photoelectric effect, ap-

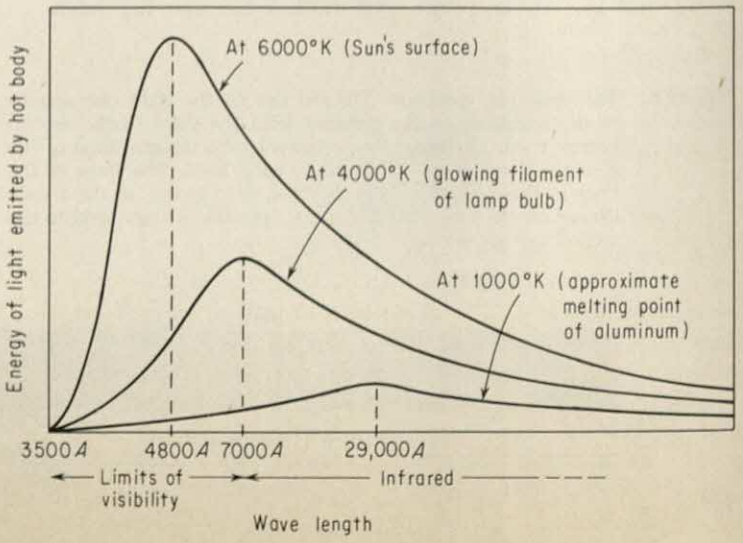


Fig. 12-6. The distribution of the energy given off by hot bodies at three different temperatures. A body at 6000°K gives off a maximum amount of energy in the visible range of the spectrum.

peared in 1905. This effect occurs when light of sufficient energy falls upon an atom and causes it to lose electrons. The energy of the light actually tears electrons away from the atoms of a substance. In Fig. 12-7, the principal parts of a simple lightmeter are shown. The energy of visible light, falling upon the thin layer of selenium, strips electrons from these atoms and deposits them on the iron by way of the external electrical circuit. This process sets up an electrical current, the amount of which can be measured by attaching an ammeter in the circuit as shown. Since the amount of electrical current is proportional to the intensity of light, this instrument indicates the amount of light present—hence, its use as a lightmeter.

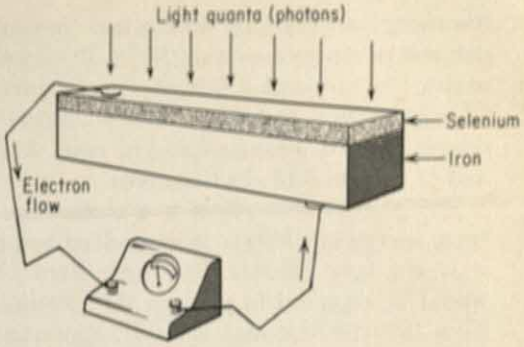


Fig. 12-7. The photoelectric effect, as employed in the photographic light meter.

In order to produce the photoelectric effect in common metals such as iron, zinc, and copper, the higher energy of ultraviolet light is required. For the alkali metals, such as sodium and cesium, however, ordinary visible light has sufficient energy to tear electrons away from their atoms. Einstein explained the photoelectric effect by concluding that light consists of light quanta, or photons, and that the energy of a photon is related to the frequency of the light, as represented by the equation $E = hn$. His result thus agreed with Planck's equation above.

The energy of the photon is used up in two ways—part of it is needed to tear the electron from the atom, and the rest is imparted to the ejected electron. Einstein expressed this transfer of energy mathematically as follows:

$$\underset{\text{energy of photon}}{hn} = \underset{\text{energy of ejection}}{E_i} + \underset{\text{kinetic energy of ejected electron}}{\tfrac{1}{2}mv^2} \qquad (12.2)$$

This relationship tells us that if a photon has enough energy to knock out an electron from an atom, the energy of the photon is used up in actually tearing the electron away from the atom and, at the same time, in imparting a certain amount of kinetic energy to it—a nice example of the law of conservation of energy. To use a mechanical analogy, if the photon is represented by a rolling bowling ball and the electron by a bowling pin, according to the law of conservation of energy, the energy of the ball knocks over the pin and, in addition, imparts kinetic energy to it.

Robert Millikan of the University of Chicago verified this equation experimentally in 1912 and obtained an accurate value for Planck's constant h. If we could find a photon that barely knocks out an electron from an atom, we would know the value for E_i, the energy of ejection. Millikan conducted experiments by which he could measure these values in terms of voltage, and we now have values for the energy of ejection of electrons for various elements. The voltage necessary to remove an electron from an atom is called the *ionization potential.* (An atom that has a charge on it—and if we knock an electron from an atom the resulting atom will be positively charged—is called an *ion.*) These ionization potentials are expressed in electron-volts. An *electron-volt* is

the energy acquired by an electron moving under a fall of one volt; it amounts to only 1.6×10^{-19} joule. The accurate measure of so small a quantity is a significant achievement of experimental technique. The ionization potential of hydrogen is 13.6 electron-volts, of neon 21.56 electron-volts, and of sodium 5.14 electron-volts.

The photoelectric effect is a method of generating electrical energy in addition to those described in Chapter 9, but it is not very efficient. The light from 20 billion candles would be required to produce the electrical energy needed for a 100-watt electrical lamp. By exposing part of a square inch of sensitive material to sunlight, scientists have succeeded in generating enough electrical energy to operate a small radio. The cost of such devices, however, is prohibitive for general use at present, although, with further study, it is reasonable to expect that they will become more and more practical. On the other hand, the photoelectric effect is even now usefully applied in lightmeters, movies, television, automatic door-openers, burglar traps, and many other devices. In each instance, the light used must provide more energy than that required for the ejection of the electron, and the quantity of electrical current produced is proportional to the intensity of the light.

How does the photoelectric effect help us in developing our picture of the atom? What is the relationship among the hydrogen spectrum, the quantum concept, and the wave motion of the electron? Bohr provided an exceptionally brilliant interpretation that has served well in guiding the physical scientist into this strange, minute world. Let us examine his view of the atom, taking into account the ideas of Planck and Einstein as well as others who later improved on Bohr's atom.

12-5 The Modern Atom

Bohr said that an excited hydrogen atom emits light in definite quantities of energy; that is, a particular line in the hydrogen spectrum represents a particular quantity of energy that can be determined from Equation 12.1. The red line in the hydrogen spectrum has a frequency of 4.5×10^{14} waves per second. From Equation 12.1, then, we can find the energy of each photon in the light in this particular line:

$$E = hn$$

$$E = 6.6 \times 10^{-34} \text{ joule sec} \times 4.5 \times 10^{14}/\text{sec}$$

$$E = 3 \times 10^{-19} \text{ joules}$$

But where does this particular quantity of energy come from? And why are there no orange or yellow lines in the hydrogen spectrum? How do we account for the green, blue, and violet lines? Bohr explained that this energy comes from the electron that is in the outer part of the atom. If the elec-

tron gives up a certain quantity of energy, it then possesses *hn* less energy. In other words, the electron in an excited atom is pushed into a higher energy level by absorbing *hn* energy from the exciting source. Then, when the electron falls back to its original position it gives off exactly as much energy as it originally absorbed because the atom is a conservative system. The clue to the amount of energy absorbed by the electron is given by the color of the light that is emitted when the electron is permitted to fall to its original position. The hydrogen spectrum then tells us exactly how much energy the electron can absorb or emit. The color lines tell us that there are different levels of energy to which the electron may climb before it falls back to its original, or ground, state.

By studying the spectrum in Fig. 12-5, as well as other line spectra in the ultraviolet and infrared regions, scientists have gained considerable insight into the hydrogen atom. The electron may occupy one of several energy levels. The lowest level is called the ground, or normal, state. The difference in energy between any two levels is given by the frequency of light that is emitted when the electron moves from the higher to the lower level (Fig. 12-8). At no time can we find an electron between these levels because its movement is like that of an elevator that stops at the various floors of a building, but never between them. More accurately, the outer part of an atom is analogous to a building with staggered elevators, that is, with one operating upward from the ground level, a second operating upward from the second floor, a third operating upward from the third floor,

Fig. 12-8. The source of the colors of light emitted by an excited hydrogen atom.

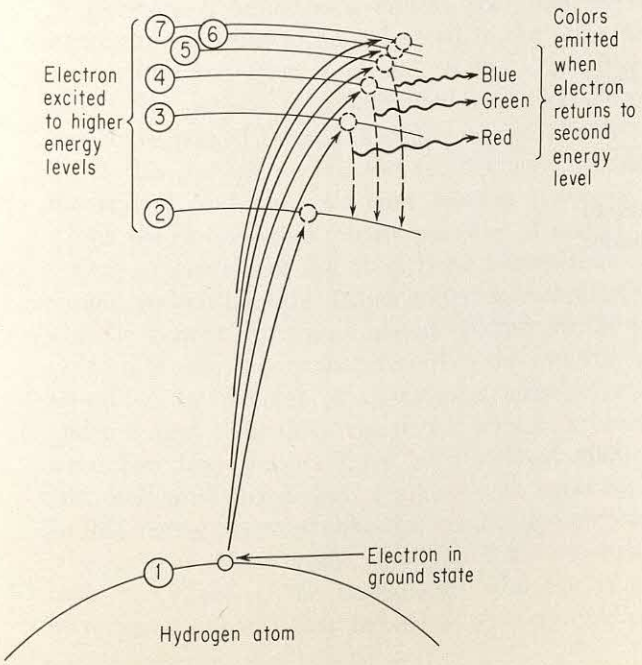
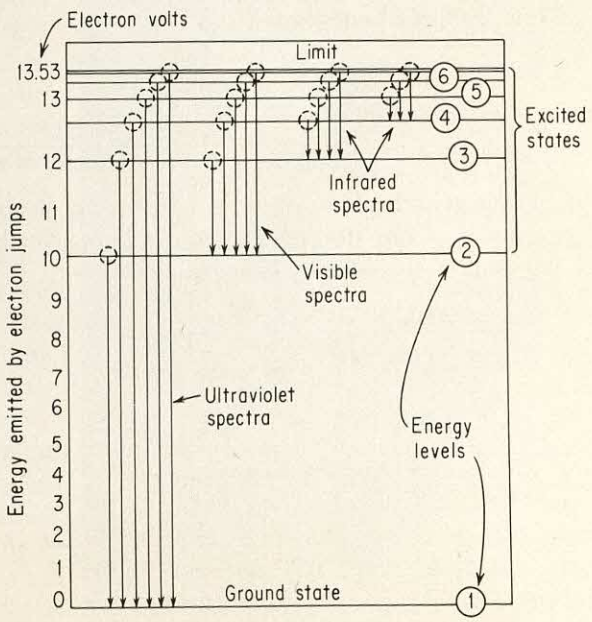


Fig. 12-9. Energy-level diagram of the hydrogen atom.



and so on. In the atom, these elevators correspond to groups of spectral lines that are found in different regions of the electromagnetic spectrum. The first elevator corresponds to a series of lines that are found in the ultraviolet region (Fig. 12-9). The second elevator corresponds to the visible lines noted in Fig. 12-5, and the higher elevators correspond to infrared line spectra.

At this point, let us stop briefly and review what we have found out about the nature of light. Light, as we saw in Chapter 10, is an electromagnetic wave. We see now, however, that, in addition to its wavelike character, it possesses energy that is proportional to its frequency. Studies of the light given off by excited atoms reveal that light may be considered as a stream of tiny bundles of energy called photons (or quanta). The photon is a particle of energy that is entirely different from any other phenomenon in science. Although it is largely a theoretical concept, best described by mathematical equations (such as Equations 12.1 and 12.2), it has supplied scientists with such a helpful explanation for the behavior of light that it is now universally accepted. The photon, however, is just one characteristic of ordinary light, for as we know, light also acts as a wave, undergoing reflection, refraction, and diffraction. This dual nature of light—its behavior as both a wave and a stream of photons at the same time—creates one of the more complex problems of science, but the dual theory is the best one available at the moment to describe the properties of light.

During the 1920's, several additions were made to our knowledge of the atom. Louis de Broglie of France found the wave length of an electron by comparing electrons to photons. He reasoned that since photons had been found to have a momentum depending upon their wave length, electrons, which we have thought of as particles up to now, should have a wave length depending upon *their* momentum. In the de Broglie equation

$$\lambda = \frac{h}{mv} \qquad (12.3)$$

$$\text{wave length of electron} = \frac{\text{Planck's constant}}{\text{Mass of electron} \times \text{velocity of electron}}$$

the wave length of an electron is inversely proportional to its velocity. Thus, for an electron whose velocity is about 1 per cent of the velocity of light:

$$h = 6.62 \times 10^{-34} \text{ joule sec}$$

$$m = 9.1 \times 10^{-31} \text{ kg}$$

$$v = 2.18 \times 10^{6} \text{ m/sec}$$

and substituting into Equation 12.3 gives the wave length of the electron:

$$= \frac{6.62 \times 10^{-34} \text{ joule sec}}{9.1 \times 10^{-31} \text{ kg} \times 2.18 \times 10^{6} \text{ m/sec}}$$

$$= 3.33 \times 10^{-10} \text{ m or } 3.33 \text{ A.}$$

This is very nearly the wave length of the electron in the ground state of the hydrogen atom.

A big step forward in understanding the structure of atoms was made in 1926 by Schroedinger and Heisenberg, who initially developed the idea of *wave mechanics*. Although the mathematics of the Schroedinger equation and its solutions are well beyond the scope of our present interest, the results of this work are important in gaining at least a rudimentary idea of the modern concept of atomic structure. The wave mechanical approach tells us that the electrons in an atom may be described by principal quantum numbers, $n = 1, 2, 3, 4, \ldots$, and these are identical to the energy levels described by Bohr. These energy levels are frequently called *orbitals* in order to distinguish the electron's behavior from that of a planet in *its* orbit.

The basic idea in wave mechanics is that matter is composed of waves. This idea is particularly evident with small particles such as electrons, and had nature been endowed with a large Planck's constant (Equation 12-1) we might have discovered the wave nature of matter much earlier. A baseball heading for the center field bleachers or a rifle bullet whizzing at 4000 ft/sec, because of their relatively large masses, have wave characteristics much too small to be significant. But an electron moving in a small box the size of an atom and with a velocity about 1 per cent of the velocity of light definitely shows wavelike characteristics.

If electrons show a wave action when they are in atoms, the next logical questions is: Do electrons *always* travel in waves? After de Broglie pointed the way, scientists soon found that a stream of electrons moves in a wave because it undergoes diffraction and interference in exactly the same way as do X rays. The structures of many substances have been studied by means of electron diffraction. When a beam of electrons is aimed at a crystal, the atoms in the crystal cause the electrons to be diffracted in the same way X rays were diffracted from a crystal in Fig. 11-6. Figure 12-10 shows the pattern of a diffracted beam of electrons after they have passed through a single crystal of clay. An analysis of this pattern reveals the arrangement of the atoms in the crystal, and thus, we have another powerful tool with which we can examine minute particles of matter.

Our conception of the atom is no longer that of Dalton's simple, indivisible unit. It has gotten so complicated that we literally cannot draw an adequate picture of the atom. The modern atomic theory that describes the electron waves surrounding the nucleus of an atom is largely mathematical, and it is very difficult to visualize a photon, or even an electron. Yet these entities have been defined mathematically, and their predicted behavior agrees with experimental facts. In this matter, we must take the word of the atomic scientist, who is an explorer fully equipped with the mathematical tools, the theories, the instruments, and the techniques to investigate a world that is beyond our comprehension. Al-

Fig. 12-10. An electron diffraction pattern of a clay crystal, proof of the wave-like character of a moving electron. (Courtesy W. O. Milligan, Rice University)

though his ideas are accurate only in equations and symbols, we must be content to approach them only in words. We must accept his story much as we accept the story of the traveler who returns from the far corners of the earth, or of the soldier who returns from a bitter campaign, or of the religious man who has talked with his God. We are somewhat in the position of a listener who hears a second-hand story that is told in a foreign language.

12-6 More Complex Atoms

Let us return to the Bohr atom, which does satisfactorily explain our wave-mechanical conception of the atom. His hydrogen atom, as we have seen, has several energy levels to accommodate the electron when it becomes excited. But what happens in an atom that has two electrons, or a dozen, or four dozen? Where would we expect to find these electrons?

The helium atom has two electrons revolving around the two protons in its nucleus, and thus the atom as a whole is electrically neutral. On our list of atomic weights, the helium atom has a mass of 4, and therefore it must contain two neutrons in its nucleus along with the two protons. A very simple diagram of the helium atom is shown in Fig. 12-11. The circle, of course, merely represents the volume occupied by the atom, and the relative sizes of the electrons, protons, and neutrons are greatly exaggerated in order to show them in the illustration.

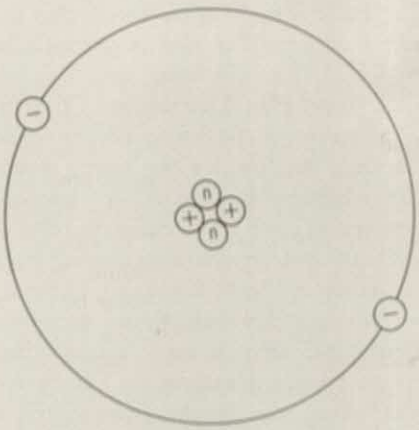


Fig. 12-11. Bohr's model of the helium atom.

A closer examination of the helium atom raises a puzzling question. Since moving electrons create a magnetic field, according to the rules discussed in Chapter 9, the two revolving electrons in a helium atom should create twice the magnetism produced by a hydrogen atom. In fact, however, the helium atom has almost no magnetism, whereas the hydrogen atom is measurably magnetic. What happens, then, to two electrons when they are paired with each other in the same energy level in the same atom? The answer is provided by a combination of facts and theory. The fact is that on an extraordinarily sensitive spectroscope, the spectral bands of an excited helium atom are composed of *two* closely spaced lines. In addition, the helium atom has no magnetic field, so that the two electrons must be cancelling each other's magnetic effect. The theory says that no two electrons in an atom can be in exactly the same state, that is, even though they are in the same energy level, they cannot be doing exactly the same thing at the same time. In other words, their wave motions are in opposite directions and they are said to have opposite spins. In the helium atom, the magnetic fields generated by the two electrons spinning in opposite directions cancel each other out, and as a result the atom is nonmagnetic. Any atom that is magnetic, on the other hand, has unpaired electrons that are not counteracted

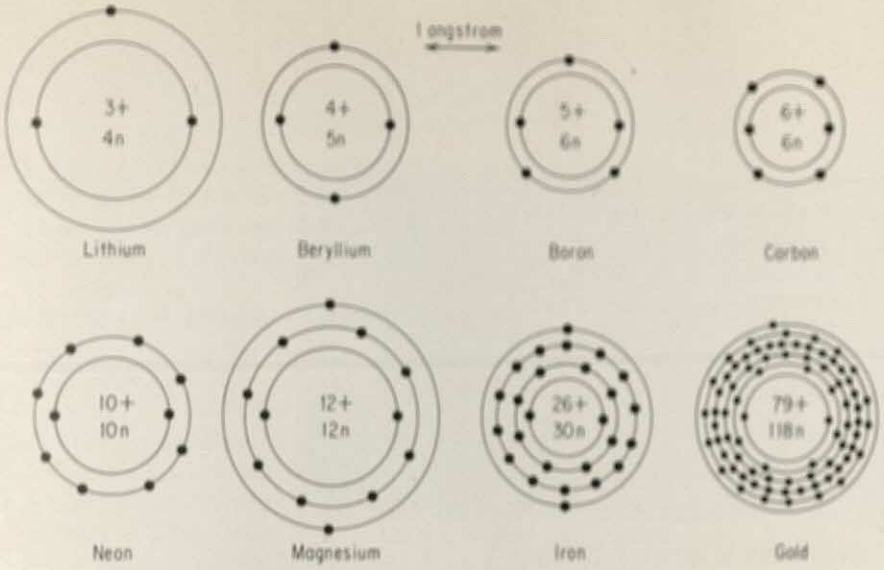


Fig. 12-12. Atomic structure of the most common isotopes of a few elements.

by other electrons spinning in the opposite direction. In the iron atom, four unpaired electrons revolve independently around the nucleus, and their combined effect produces one of the most highly magnetic of atoms.

According to the quantum theory, only two electrons can exist in the first energy level. In an unexcited atom containing three electrons, such as lithium, two electrons are in the ground level, and the third must be in the second level (Fig. 12-12). The number of electrons in these rings is determined by rather simple rules. The second energy level can accommodate up to eight electrons, that is, there are apparently enough energy states within the second energy level to permit four pairs, or eight electrons. The third level can accommodate up to 18 electrons, the fourth, 32 electrons. The higher levels undoubtedly can accommodate still more electrons, but there are no known elements that have enough for us to find out.

The diagrams of atoms in Fig. 12-12 show only the electron structure in unexcited atoms. Since we have found from the excited hydrogen atom the difference in energy between the energy levels in the atom, we will now be concerned mostly with atoms under normal conditions. All atoms potentially have all the energy levels found in the hydrogen atom, but these levels will be occupied only to the extent that there are electrons to fill them. We may think of atoms as hotels in which the floors are the energy levels, with automatic elevators connecting them, and the guests are the electrons. The hotel has floors whether or not there are guests to occupy all of them. We studied the hydrogen atom in detail by sending *one* guest up to explore the rooms in the upper floors. Now we shall see what happens when more guests arrive. If a certain number of electrons, protons, and neutrons were, by chance, to form an atom, the electrons would trickle down from the highest energy levels and fill up the lowest energy levels. The number of electrons in

TABLE 12-2 ARRANGEMENT OF ELECTRONS IN THE NOBLE GASES

	Number of electrons in each energy level						
	1	2	3	4	5	6	7
Helium	2						
Neon	2	8					
Argon	2	8	8				
Krypton	2	8	18	8			
Xenon	2	8	18	18	8		
Radon	2	8	18	32	18	8	

these various levels for the noble gases is shown in Table 12-2. Note that the outermost level in any atom never contains more than 8 electrons, and the next to outermost level never more than 18 electrons.

Let us take a backward look for a moment. We now have a theory that explains the atomic structure of a hundred different elements. They are all made up of electrons, protons, and neutrons, and the electrons occupy certain energy levels. It would appear that these elements, all composed of the same building blocks, would be similar to each other, but as we shall learn in the next few chapters these elements differ greatly. Much of the difference between atoms is caused by the number of electrons that occupy a particular energy level. For instance, an atom with one electron in, say, the third energy level is vastly different from an atom that has two or three electrons occupying the third level.

There are atoms that never exist in nature by themselves, atoms that are good electrical conductors, atoms that form extremely hard materials and extremely soft materials, atoms that react violently with each other, and atoms whose behavior is drab and uninteresting. In all these instances, however, the atoms are made up of electrons, protons, and neutrons. If we know how these basic components of the atom are arranged and how they function, we will come closer to explaining why the atom is what it is.

12-7 Atomic Numbers

Up to now we have found several good methods of comparing the weights of atoms of different elements to each other. This comparison is valuable to the extent that a system of atomic weights has been set up. More important to the structure of the atom, however, is the concept of atomic number. The *atomic number* of an element is the number of protons in its atom. How these numbers were arrived at is a dramatic and significant chapter in physical science. In his "shotgun" experiments, Rutherford found that alpha particles are deflected more by heavier atoms than by lighter ones. This fact he attributed to the difference in positive

charge that existed on the nucleus of the atom—the heavier the atom the greater the positive charge on the nucleus. Although these early results were not very accurate, they pointed toward the idea that there is a positive charge on the nucleus.

To H. G. J. Moseley goes the credit for discovering the way to determine the exact charge on a nucleus. This brilliant, young English scientist, using an X-ray tube (see Fig. 10-26), examined the wave length of the X rays that are produced when electrons strike a metal target. The stream of electrons, traveling at about one-tenth the velocity of light, penetrates the atoms of the metal and collides with an electron in the inner level of an atom. The high energy of the incoming electrons ejects an electron from the inner level (Fig. 12-13), and the resulting vacancy is immediately filled by an electron that jumps in from an outer level. In the process, this electron gives off energy in the form of an X ray.

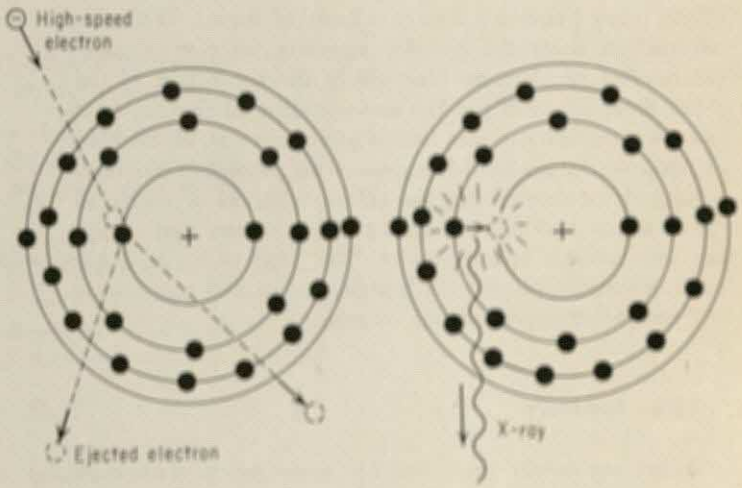


Fig. 12-13. A high-speed electron penetrating the outer parts of an iron atom collides with an electron in the innermost energy level.

Moseley examined the wave length of the X rays produced when the target metal was iron, copper, zinc, and so on. He found a gradual, but definite, decrease in wave length as the atomic weight of the metal increased. The Moseley relationship is shown in Fig. 12-14. After much experimentation, he concluded that the wave length of the X rays is determined by the total number of electrons in the atom. He discovered that as the number of electrons increased the wave length decreased, and, conversely, that the greater the number of electrons the greater was the frequency of the emitted X ray. Mathematically, the Moseley relationship is stated:

$$\sqrt{n} = aZ - b$$

In this equation, $\sqrt{n}$ is the square root of the frequency, a and b are numerical constants, and Z is the number of electrons in the atom.

Fig. 13-16 The wave lengths of the X rays produced by an X-ray tube with various metallic targets

If we wish to determine the atomic number of an element, we make this element the target in the X-ray tube. Then the wave length of the X rays is determined by reflecting them from a crystal of known structure. From the wave length of the X rays, their frequency can be found (wave length times frequency equals velocity of wave). With this information, using the Moseley equation, we can calculate the number of electrons that are in the outer part of the atom. Since an atom has the same number of protons as it has electrons, we can also immediately arrive at the number of positive charges on the nucleus. This procedure has been employed for determining the atomic numbers of nearly all the elements and for identifying new elements and verifying their atomic characteristics. This kind of analysis has been extremely valuable in bringing order and meaning to our comprehension of atomic structure.

13-8 Summary

Atoms are mostly space and are composed of three main parts: electrons, protons, and neutrons. The protons and neutrons are confined to the nucleus, which is about 1/10,000 the diameter of the atom, and the electrons occupy energy levels in the outer parts of the atom. These electrons can be excited to higher energy levels, and when they fall back to their original level they emit light of distinctive wave lengths.

The Bohr atom, although it does not fulfill all the requirements for an ideal atomic theory, is a useful and productive concept. It explains the spectra of excited atoms and agrees nicely with Planck's quantum theory of the photon. The photon, a unit quantity of energy, is substantiated by the photoelectric effect which was explained by Einstein.

De Broglie showed that electrons, like photons, have wave-like properties related to their energy and, consequently, produce electron diffraction.

In the atom, no two electrons can occupy the same energy level unless their spins are in opposite direction. Atoms of

all elements differ, even though they are composed of the same basic building blocks.

More complex atoms contain K, L, M, and N electrons in successive energy levels. The atomic number of an element is the number of protons in the atom.

EXERCISES

A. KEY TERMS TO REMEMBER

alpha ray	energy level	photoelectric effect
atomic number	gamma ray	photon
beta ray	ground state	Planck's constant
helium atom	hydrogen spectrum	quantum theory
electron cloud	light quanta	nucleus

B. QUESTIONS ABOUT ATOMIC STRUCTURE

1. Discuss the parallelism in the statements: atoms are to matter as photons are to energy.

2. What kind of particles besides alpha particles could be used in Rutherford's experiment in which he bombarded gold atoms?

3. By a diagram, show the wave motion of an electron in the ground state of a hydrogen atom. Then show the wave motion in the second energy level. Now explain why an electron cannot exist between these two levels.

4. Explain the similarities and differences between the solar system and the atom.

5. Tell how the discovery of the photoelectric effect led scientists to modify their view of light as an electromagnetic wave.

6. What proof do we have that electrons, when in motion, have wavelike characteristics?

7. Draw a simplified picture of the electron structure of the following atoms, the atomic numbers of which are given in parentheses: helium (2), beryllium (4), carbon (6), fluorine (9), neon (10), sodium (11), aluminum (13).

8. If the atom is apparently mostly space occupied by a small nucleus and the wave action of electrons, why should it be so difficult to penetrate the atom? Can one atom be made to penetrate another atom?

9. According to Fig. 12-8, what is the most intense color in the light from the sun. Does this suggest a connection between the sun's light and the predominant color we see among growing plants?

C. PROBLEMS ABOUT ATOMIC STRUCTURE

1. If an electron has a velocity of 1×10^8 cm/sec, what is its wave length as determined by the de Broglie equation?

2. The maximum wave length that will knock an electron off a magnesium atom is 3700 Å. Change the wave length to the

quency, calculate the energy of ejection for his photoelectric effect, and express this energy in electron-volts. Ans. 5.4×10^{-19} joules or 3.34 electron-volts.

3. What is the wave length and frequency of radiation whose energy is equal to 4×10^{-20} joule per photon?

4. Using de Broglie's equation, find the wave length of (a) a golf ball (44 grams) traveling with a velocity of 100 m/sec and (b) an electron traveling at one tenth the velocity of light. Why is there a great difference in these two wave lengths?

Thirteen

The Elements

Every common substance we know of is composed of one or more of about sixty elements, either as simple elementary substances or in combinations called compounds. Additional elements have been found in very small traces, or they have been synthesized in small quantities to bring the total number of elements on the earth to slightly over a hundred. Do additional elements exist elsewhere in the universe? We do not know for sure. Astronomical studies tell us that some of the stars are composed of something far denser than anything we know on earth. These dense stars may be composed of compressed or collapsed atoms, although it is more likely that they are composed of neutrons, protons, electrons, and other particles even more elementary than our so-called elements. This poses the difficulty of defining the term "element." There is evidence that some stars contain short-lived elements with an inside-out structure—that is, elements whose atoms have a negatively charged nucleus and positively charged electrons or positrons. On the other hand, the spectra of the stars, including our sun, show that many elements familiar to us—calcium, carbon, hydrogen, and oxygen—occur in the remote parts of the universe. Furthermore, the analysis of meteorites that arrive to us from space has never shown other elements than those familiar to all of us, such as iron, nickel, and silicon.

In our present study, we will concern ourselves primarily with the elements that exist within the scope of our direct experience. In cosmic terms and, indeed, even relative to the size of the earth, this scope is quite limited. We can analyze accurately the atmosphere up to a height of several miles, the earth's crust down to a depth of a few miles, and most of the water on the earth, but beyond these limits we

must rely upon indirect, and often inconclusive, evidence for our information. In Chapter 10 we saw how earthquake waves tell us that the earth has a core of great density, indicating that it is probably a mixture of iron and nickel. These waves also reveal that the mantle is composed of a more complex, but less dense material than the core. Our inability to obtain direct evidence of the composition of the core and the mantle, however, drastically restricts our knowledge of their elementary composition.

13-1 Discovery of the Elements

Prior to the seventeenth century little distinction between elements and compounds was made. Many of the elements we know of today—sulfur, iron, copper, silver, gold, etc.—have been produced and used for thousands of years. But the basic concept of an element as a simple substance has evolved from rather primitive beginnings. The Greek philosopher-scientists of the fifth century B.C. proposed that all substances were somehow related to the four "elements"—earth, air, fire, and water. Much later, during the Middle Ages, considerable effort was made to convert certain common metals to precious ones, an idea possibly suggested by the success of those who converted ores such as red iron oxide to a white, useful metal. The search for the "philosopher's stone" and the "elixir of life" during this same period was a failure, although certain information, coming as by-products of this activity, later became useful in developing a more useful concept of an element.

Robert Boyle is credited for giving a specific meaning to the term "element" in 1661, and from this meaning has developed our modern concept of an element. By the time of the American Revolution, the word was generally accepted as referring to those substances that could not be separated into simpler substances. Thus, a distinction was established between the simpler or elementary substances such as iron, gold, and sulfur, and the more complex substances we call compounds. In 1789, about thirty elements had been identified, and at the outbreak of the Civil War, over sixty elements were known. Partly as the results of a more useful concept of an element and a greater concerted effort to experiment scientifically, more elements were discovered in the last century than during the entire period of history prior to 1800.

The symbols of many elements reveal the time of the discovery of the element. For example, the following elements whose symbols were derived from Latin were known to the Romans: iron (Fe, from *ferrum*); gold (Au, from *aurum*); silver (Ag, from *argentum*); lead (Pb, from *plumbum*); tin (Sn, from *stannum*); copper (Cu, from *cuprum*); and mercury (Hg, from *hydrargyrum*—quicksilver). Although the symbols for sodium (Na, natrium) and potassium (k,

kalium) were derived from Latin words, these elements were not discovered until 1807. The symbols of most of the elements, however, are based on the present names of the elements.

Many elements were named after their dominant characteristic. Astatine (At) is unstable (*astatic*) and must be synthesized in the laboratory. Osmium is a metal that forms an oxide with a strong odor (from the Greek word *osme*, odor). Promethium, for a time believed to be impossible to detect, was for this reason named after the Greek god, Prometheus, who stole fire from heaven; it evaded all efforts to discover it and eventually had to be synthesized in the laboratory. Chromium (from the Greek *chroma*, color) forms highly colored compounds that are widely used for pigments. Nickel, because its ore appeared to contain copper but yielded none, was called "kupfernickel" after the German word for "copper" and "demon." Cobalt refuses to be easily separated from nickel and therefore was called a "goblin" (German *kobalt*, goblin). Since argon, a gas that composes almost 1 per cent of the atmosphere, never reacts with other elements, its name was derived from the Greek word *argon*, meaning inactive. Before being found on the earth, helium was detected in the sun's atmosphere through an examination of its spectrum; its name was therefore derived from the Greek word *helios*, meaning sun.

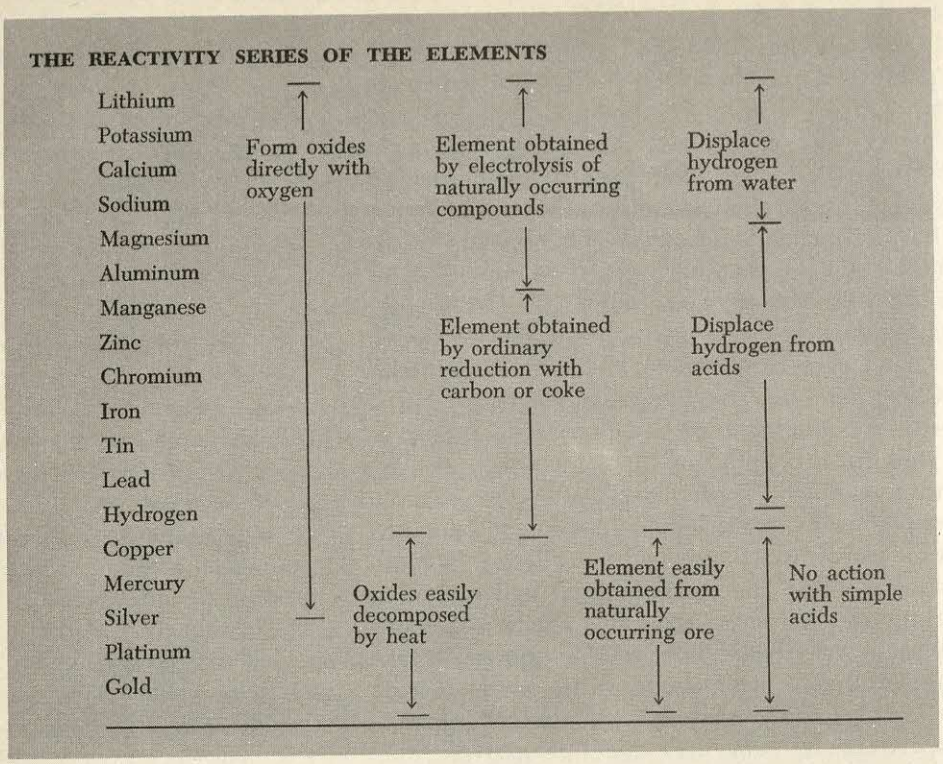
Other elements were named after their discoverer or the country in which they occur or were produced. By the beginning of the twentieth century, most of the naturally occurring elements had been discovered. The rate of discovery of new elements was speeded up during the nineteenth century by the growing interest in science and by the development of successful methods of predicting the properties of unknown elements.

13-2 The Reactivity Series of the Elements

If we examine a list of the elements and the compounds in which they occur, we will find that some of them occur naturally as free elements and others exist only in compounds. Why do we find free, metallic gold occurring in nature, but never any free sodium, even though nearly 3 per cent by weight of the ocean is sodium chloride? The answer is based upon one of the most fundamental characteristics of the elements—their reactivity. The *reactivity* of an element is its tendency to form compounds. The most reactive elements occur naturally only in compounds; the least reactive elements occur as free elements in nature. Elements of intermediate reactivity may occur either free or combined.

If we arrange a list of the elements with the most reactive ones at the top and the least reactive ones at the bottom, we have the so-called reactivity series of the elements (Table

TABLE 13-1



13-1). From it we can make several interesting deductions. Those elements at the bottom (beginning with gold up) have been known since prehistoric times; with the exception of hydrogen, the bottom eight were known to the Romans, who had the means of obtaining these metals from their ores and thus were able to use them in their vehicles, utensils, buildings, and artifacts. The great Roman baths were supplied with water by lead pipes, hence the source of our word "plumbing" (*plumbum* from the Latin for lead). Going up in the series, the elements become more and more difficult to separate from their ores, and only by the most vigorous of processes, electrolysis, are those elements at the top released from their compounds.

The comparison, however, between the reactivity series and the natural occurrence of the elements is not parallel in all cases. Although sulfur occurs frequently as a free element in nature, it does so only if it is protected from long exposure to the atmosphere. Oxygen, a reactive element, occurs free in the atmosphere only because it is steadily replenished by the action of photosynthesis. The reactivity series in Table 13-1, consequently, includes only those elements that can be classed according to their inherent reactivity.

13-3 The Natural Occurrence of the Elements

Fortunately for us, certain elements are concentrated in ore deposits. Little is known about how these deposits were

formed, but from them we can extract many elements that might otherwise be unavailable to us because they would be scattered throughout the earth's crust, diluted by the more common elements. Helium, for instance, is an extremely rare element, but it has accumulated in certain gas wells in Texas, Oklahoma, and Kansas to the extent that it can be commercially separated and used to fill balloons and to prevent the bends in deep-sea divers. Iron, on the other hand, is plentiful in the earth's crust, but our modern civilization depends upon it being concentrated in large, relatively pure deposits. The iron ores of Minnesota, containing upward of 50 per cent iron, have been mined continuously for many years. Although aluminum is the most abundant metal in the earth's crust, the most useful ore is a certain oxide, bauxite, which has been collected by geochemical processes into convenient deposits in Arkansas and the Guianas where it may be literally scooped up from the earth's surface and hauled away.

Let us consider briefly some of the processes that occur naturally to produce concentrations of certain elements in useful ore deposits. Ore deposits are the result of geological processes that sort out or concentrate certain elements. The exact nature of these processes is not always well understood, although some of them can be subjected to laboratory study and have been verified. Sedimentation, which is the dropping of dissolved or suspended particles from solution, accounts for the formation of many ores. A wide variety of salts, including ordinary rock salt and gypsum, as well as iron ores and gold deposits, are found in various parts of the world in sedimentary layers. At other times, ore deposits are formed through selective dissolving. That is, if a part of the earth's crust contains a mixture of substances that vary in solubility, the most soluble components will be eroded away first, leaving behind the least soluble components. These least soluble components—the residual deposits—may consist of iron ores such as those in Minnesota and the bauxite ores in Arkansas.

Down inside the earth's crust there is yet another important process of ore formation that may gather certain elements into economically usable deposits. Here, the temperatures are high enough actually to dissolve appreciable amounts of metal compounds. These solutions trickle or percolate through cracks and breaks in the rock, especially along old fault lines. As these hot solutions move upward to cooler zones, the metallic substances come out of solution and are deposited along the old fault lines. Recurrent faulting along these lines opens them up for more deposition of ores, until a deposit known as a *lode* or *vein* is formed. Such veins, rich in copper, lead, zinc, silver and gold, are eagerly sought by the prospector and may sometimes be several feet thick and extend hundreds of feet along an old fault line.

A list of the more common elements, their more common

TABLE 13-2 SOME MINERALS AND METHODS OF OBTAINING METALS FROM THEIR ORES

Metal	*Minerals*	*Method of obtaining metal*
Aluminum	clays, feldspar, bauxite Al_2O_3	electrolysis
Calcium	gypsum, chalk, limestone $CaCO_3$	electrolysis
Chromium	chromite $Fe(CrO_2)_2$	reduction with coke
Copper	chalcocite Cu_2S, cuprite Cu_2O	reduction with coke
Gold	free element	mechanical separation
Hydrogen	water H_2O	reduction with coke
Iron	pyrite, magnetite, hematite Fe_2O_3	reduction with coke
Lead	galena PbS	reduction with coke
Lithium	spodumene $LiAl(SiO_3)_2$	electrolysis
Magnesium	magnesite $MgCO_3$, in sea water as $MgCl_2$	electrolysis
Mercury	cinnabar HgS	roasting
Platinum	free element	mechanical separation
Potassium	sylvite KCl	electrolysis
Silver	free element, sulfide Ag_2S	roasting
Sodium	common salt NaCl	electrolysis
Tin	cassiterite SnO_2	reduction with coke
Zinc	sphalerite ZnS	roasting, reduction with coke

forms of natural occurrences, the chemical composition of their ores, and the method of obtaining them from their ores is given in Table 13-2. From this table, we can see that only by the most vigorous method (electrolysis) are the most reactive elements obtained from their ores. The separation methods in order of complexity are: electrolysis, reduction with coke, roasting, and mechanical separation. If we arrange these elements according to the difficulty encountered in obtaining the free element from the ore, we will have a list very similar to that of the reactivity series in Table 13-1.

13-4 The Abundance of the Elements

Let us turn to the average composition of those portions of the earth that have been fairly well analyzed. The atmosphere over a square mile of the earth's surface has a mass of over 30 million tons. Although this seems like an enormous weight, the atmosphere of the whole world, composed chiefly of nitrogen and oxygen, is less than one-millionth the mass of the earth. At sea level its average composition, expressed in terms of the relative number of atoms of each element, is given in Table 13-3. In this table, we see that in each thousand atoms of the atmosphere there are 780.3 nitrogen atoms. The same method is used below to express the composition of the ocean and the earth's crust.

The nine most plentiful elements in the ocean are shown in Table 13-4. Almost all the hydrogen and oxygen in the ocean are combined as water. Most of the other elements in sea water are present in the form of soluble salts. More than forty elements have been found in sea water, which contains an almost inexhaustible source of many valuable

TABLE 13-3 COMPOSITION OF DRY AIR

	In Number of Atoms per Thousand Atoms of Atmosphere
Nitrogen	780.3
Oxygen	209.9
Argon	9.3
Carbon dioxide	0.35
Hydrogen	0.1

TABLE 13-4 AVERAGE COMPOSITION OF THE OCEANS

	In Number of Atoms per Thousand Atoms of Sea Water
Hydrogen	661.06
Oxygen	330.53
Chlorine	3.35
Sodium	2.87
Magnesium	1.71
Sulfur	0.33
Calcium	0.06
Potassium	0.06
Bromine	0.005

elements. Magnesium and bromine are already being profitably taken from the sea. The other elements must await the development of more successful methods to win them from the extremely dilute solution in which they occur. Practically all of the calcium and silica that reach the ocean through rivers are precipitated or deposited by marine organisms and are thus converted into solid rock. In the geological past, huge deposits of calcium carbonate have been formed in this way into coral reefs. There is evidence that the large iron ore deposits in the Lake Superior region may have been formed by microscopic bacteria that fed on the iron in ocean water and then left behind an iron oxide deposit that was eventually compacted into iron ore.

The solid portion of the earth's crust that extends downward a few miles is called the *lithosphere*. Its composition has been analyzed by various methods; the results give the composition listed in Table 13-5.

Oxygen is by far the most abundant element, although most of it is combined with other elements in the form of rocks and minerals. The second most abundant element, silicon, has been seen by relatively few people in a free form; it is always combined with other elements in such substances as quartz or in complex silicates that include granite, felspar, mica, kaolin, and clays of many kinds. Aluminum, number 3 in rank, occurs in granite, feldspar, and clays, but it is separated industrially from an oxide, bauxite, by an electrolytic process developed by Charles Hall in 1886. It is interesting to note that some of the most familiar elements such as copper, tin, zinc, and lead, are not among the

TABLE 13-5 COMPOSITION OF THE EARTH'S CRUST

In Number of Atoms per Thousand Atoms of Lithosphere

Rank	Element	Atomic number	Atoms per thousand
1	Oxygen	8	605.0
2	Silicon	14	204.5
3	Aluminum	13	62.5
4	Hydrogen	1	27.0
5	Sodium	11	25.8
6	Calcium	20	18.9
7	Iron	26	18.7
8	Magnesium	12	17.9
9	Potassium	19	13.8
10	Titanium	22	2.7
11	Phosphorus	15	0.86
12	Carbon	5	0.55
13	Manganese	25	0.38
14	Sulfur	16	0.34
15	Fluorine	9	0.33
16	Chlorine	17	0.28
17	Chromium	24	0.15
18	Barium	56	0.08
19	Nickel	28	0.07
20	Vanadium	23	0.07

twenty most abundant elements. Indeed, titanium, a metal that until now was very difficult to obtain in useful quantities, is a hundred times more abundant than copper.

A significant aspect of these three tables is the fact that the most abundant elements have low atomic numbers. The elements above atomic number 26 (iron) account for less than 0.2 per cent of the earth's crust, ocean, and atmosphere. These substances, in short, are made up almost entirely of a small number of the lighter elements. To explain this situation, scientists presume either that the prevailing conditions were unfavorable to the heavier elements when the elements were formed, or that the elements have undergone decay into lighter elements during the course of time.

A further examination of the plentiful elements reveals that the even-numbered elements are more abundant than the odd-numbered ones. Figure 13-1, which gives the abun-

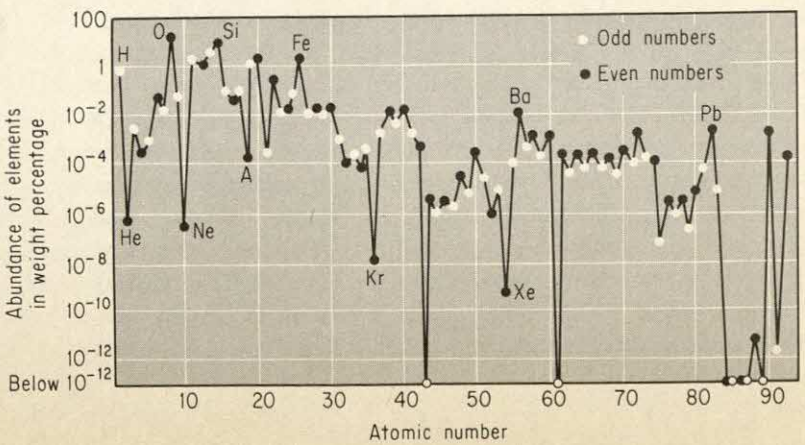


Fig. 13-1. The relationship between the abundance of the element and their atomic numbers.

dances of all the known elements that occur in nature, indicates that atoms with an even number of protons in their nuclei are more common than those having an odd number. The number of neutrons in the nucleus of the atom, moreover, apparently influences the abundance of a particular element's isotope. An even-even proton-neutron count in the nucleus is much more common than an even-odd proton-neutron count. An odd-odd proton-neutron count is the rarest kind of nucleus to occur naturally. In the 37 most abundant isotopes of the elements in the earth's crust, the proton-neutron ratio, indicated below by the number of atoms per thousand atoms of lithosphere, is overwhelmingly even-even:

even-even proton-neutron count	856.3
odd-even proton-neutron count	131.1
even-odd proton-neutron count	12.4
odd-odd proton-neutron count	none

These relationships again focus our attention on the mysterious forces that hold the nucleus of the atom together. Why should an even number of protons and an even number of neutrons get along with each other better than an odd number of these particles? The answer is simple; we do not know.

As we learned in the last chapter, the laws of motion, of gravitational attraction, and of electrical attraction and repulsion do not seem to fit the atom. How, for instance, can two protons and two neutrons exist together in the nucleus, whose diameter is less than 10^{-12} cm? Coulomb's law, derived from experiments conducted at ordinary, measurable distances in the laboratory, states that two like charges will repel each other with a force that is inversely proportional to the square of their distance apart. When the distance between these two positive charges becomes excessively small, however, this law is no longer valid. At 10^{-12} cm (Fig. 13-2), the repulsion between the two positive charges is somehow transformed into a strong force of attraction. Whether we will have to revise our present laws or formulate a new set to explain such deviations in the expected behavior of microcosmic matter is one of the unsolved problems of science. We shall examine this problem again in Chapter 26 when we consider the source of the tremendous energy locked in the nucleus of the atom.

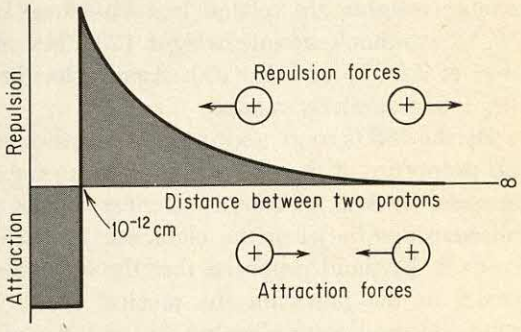


Fig. 13-2. The force of repulsion predicted by Coulomb's law does not fit the case when two protons approach each other at a distance of 10^{-12} cm in the nucleus of the atom.

Turning briefly to the composition of the sun, the planets, and the stars, we know through spectra studies that many of the elements of the earth occur in these bodies. The sun's spectrum indicates that over sixty of the earth's elements are in the sun's atmosphere, the most abundant being hydrogen (almost 82 per cent), with helium second (about 18 per cent); only very small amounts of the other elements have been detected. The composition of many of the other stars is similar to that of the sun.

Although astronomers can study only the atmospheres and the reflected light of the inner planets, they believe that these

planets are composed of elements similar to those on the earth. Venus has an atmosphere containing large amounts of carbon, hydrogen, and oxygen. Mars has a relatively thin atmosphere containing a considerable quantity of carbon dioxide, and its surface may contain iron, oxygen, and hydrogen in the forms of iron oxide and water. The giant planets are composed of large amounts of nitrogen, carbon, and hydrogen. Since their densities are similar to that of water, their composition differs considerably from that of the earth. There is evidence that the moon, Mercury, and Pluto are made up of materials that differ little from the earth's composition.

13-5 The Periodic Classification of the Elements

The periodic classification of elements was devised by scientists who noted similarities among various elements. We have all noticed that copper and gold are yellow and that silver has a slight yellowish cast. More important, these three elements are similar to each other in usage—they are our coinage metals and are used widely in jewelry. Tin and lead are two other elements that are obviously similar in appearance—they are both soft, malleable metals. These are but a few of the groups that show definite similarities in appearance and usage. As early as 1829, Johann W. Döbereiner in Germany noted groups of similar elements whose atomic weights are related to each other. In the group Li, Na, K, sodium's atomic weight (23) lies midway between that of Li (7) and K (39). Again, the three halogens Cl, Br, I, are similarly related.

By the 1860's more and more information about the chemical properties of the elements was being found, and with the increase in data, of course, scientists could more easily spot relationships between the elements. In 1866, J. A. R. Newlands in England proposed that the elements could be compared to the notes on the musical scale. In one line, he arranged the lighter elements in the order of their increasing atomic weights (noted on second line), and under this listing

Li	Be	B	C	N	O	F	Na	Mg	Al	Si	P	S	Cl	K
7	9	11	12	14	16	19	23	24	27	28	31	32	35.5	39
Do	Re	Mi	Fa	So	La	Ti	Do	Re	Mi	Fa	So	La	Ti	Do

he placed the notes of the musical scale: He reasoned that since Li, Na, and K (as well as B and Al, C and Si, and so on) are chemically very similar, they harmonize with each other as do musical notes an octave apart. From the parallel between these scales, he argued that there might be a valid relationship between the elements and music. Newlands' "law of octaves" did not include the heavier elements and it was generally ridiculed by his fellow scientists. We now look upon his arrangement as pure coincidence, though it probably served to stimulate later scientists in a search for a more basic relationship.

Series	Group I	II	III	IV	V	VI	VII	VIII
1	H							
2	Li	Be	B	C	N	O	F	
3	Na	Mg	Al	Si	P	S	Cl	
4	K	Ca		Ti	V	Cr	Mn	Fe Co Ni
5	Cu	Zn			As	Se	Br	
6	Rb	Sr	Y	Zr	Nb	Mo		Ru Rh Pd
7	Ag	Cd	In	Sn	Sb	Te	I	
8	Cs	Ba	Di	Ce				
9								
10			Er	La	Ta	W		Os Ir Pt
11	Au	Hg	Tl	Pb	Bi			
12				Th	u			

Fig. 13-3. Periodic classification of the elements according to Mendeleev, 1872.

In 1868, the Russian chemist, Dmitri Mendeleev, arranged the elements in a tabular form that is essentially the one still used today (Fig. 13-3). He placed the elements in rows according to increasing atomic weights and repeated the rows so that similar elements fell into columns. This arrangement included the similarities suggested earlier by Döbereiner and Newlands, and greatly extended the idea to include all of the nearly sixty elements known at that time. It is noted that several gaps appear in Mendeleev's table, but he was so certain of having discovered a fundamental relationship that he predicted the properties of elements that, up to that time, had not been discovered. Even though his table was not immediately accepted, he had the satisfaction of living to see several of these predicted elements discovered, with properties that agree well with the ones he forecast. Table 13-6 lists the properties of element 32, now called germanium, as predicted by Mendeleev and as they were actually discovered. Because his predictions were largely borne out, his table of the elements became universally accepted.

Mendeleev's table is a marvel of scientific organization. Although his table has had to be modified somewhat in order to accommodate all the elements and to overcome certain minor difficulties, our modern periodic table (Fig. 13-4) is based directly upon his original one. The modern table consists primarily of an arrangement of the elements in the order

TABLE 13-6 **PREDICTED AND DISCOVERED PROPERTIES OF ELEMENT 32**

	Predicted by Mendeleev	*Discovered*
Atomic weight	72	72.6
Density	5.5 g/cm^3	5.35 g/cm^3
Color	gray	grayish-white
Formula of oxide	XO_2	GeO_2
Density of oxide	4.7 g/cm^3	4.703 g/cm^3
Formula of chloride	XCl_4	$GeCl_4$
Density of chloride	1.9 g/cm^3	1.887 g/cm^3

Periods	IA	IIA						— Groups —					IIIA	IVA	VA	VIA	VIIA	0
1	1 H 1.008	← Light metals											Non-metals				Inert gases →	2 He 4.003
2	3 Li 6.94	4 Be 9.01						Heavy metals					5 B 10.8	6 C 12.0	7 N 14.0	8 O 16.0	9 F 19.0	10 Ne 20.2
3	11 Na 23.0	12 Mg 24.3	IIIB	IVB	VB	VIB	VIIB	VIIIB			IB	IIB	13 Al 27.0	14 Si 28.1	15 P 31.0	16 S 32.1	17 Cl 35.5	18 Ar 39.9
4	19 K 39.1	20 Ca 40.1	21 Sc 45.1	22 Ti 47.9	23 V 50.9	24 Cr 52.0	25 Mn 54.9	26 Fe 55.8	27 Co 58.9	28 Ni 58.7	29 Cu 63.5	30 Zn 65.4	31 Ga 69.7	32 Ge 72.6	33 As 74.9	34 Se 78.9	35 Br 79.9	36 Kr 83.7
5	37 Rb 85.5	38 Sr 87.6	39 Y 88.9	40 Zr 91.2	41 Nb 92.9	42 Mo 95.9	43 Tc 99	44 Ru 101.7	45 Rh 102.9	46 Pd 106.7	47 Ag 107.9	48 Cd 112.4	49 In 114.8	50 Sn 118.7	51 Sb 121.8	52 Te 127.6	53 I 126.9	54 Xe 131.3
6	55 Cs 132.9	56 Ba 137.4	57-71 see below	72 Hf 178.6	73 Ta 180.9	74 W 183.9	75 Re 186.3	76 Os 190.2	77 Ir 193.1	78 Pt 195.2	79 Au 197.2	80 Hg 200.6	81 Tl 204.4	82 Pb 207.2	83 Bi 209.0	84 Po 210	85 At 211	86 Rn 222
7	87 Fr 223	88 Ra 226.1	89-103 see below															

Lanthanide series	57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu
Actinide series	89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lw

Fig. 13-4. The modern form of the periodic table of the elements.

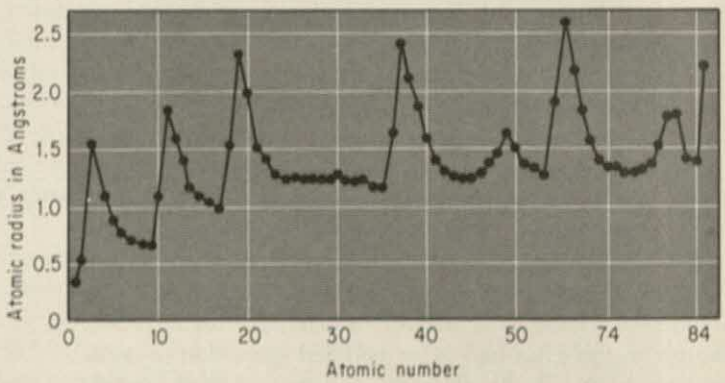


Fig. 13-5. The periodic changes of the atomic radii of the elements as a function of their atomic numbers.

of increasing atomic numbers. The elements are placed in horizontal rows, or *periods*, and at recurrent intervals elements with similar properties appear. In the first period, this interval contains two elements, the second and third periods each contain eight elements, the fourth contains eighteen, and so on. Those elements that are similar fall into vertical columns, or *groups*. There are seven periods and eight major groups in the table, and the groups may be subdivided into eighteen subgroups.

From this arrangement is derived the *periodic law*, which says that *the properties of the elements are periodic functions of their atomic numbers*. In other words, properties, such as atomic radius, hardness, density, melting point, and reactivity undergo periodic ups and downs. From Fig 13-5 you can see that these changes are amazingly regular.

13-6 Uses of the Periodic Table

The fact that the periodic table (Fig. 13-4) hangs on the walls of science classrooms and laboratories throughout the world testifies to its value. Once we become acquainted with the reasons why the elements fall into the periods and groups that they do, we will see why the table is so useful. Let us consider several of the more important uses of the periodic table, bearing in mind that there have been countless times when individual scientists have turned to the table to seek an answer to a problem, or to find a hint of some subtlety of the elements that had never occurred to them before.

Predicting the existence, and the properties, of unknown elements was one of the first uses to which the periodic table was put. Suppose we trace how Mendeleev predicted element 32. In the table he noted the elements immediately above, below, and to either side of that element. Since in other parts of the table an element's atomic weight lies almost midway between that of its neighbors, he calculated that element 32 should have an atomic weight midway between 28 and 118.7, and between 69.7 and 74.9. Hence, he predicted 72 for its atomic weight. He knew that the elements in Group IVA form oxides with the formulas CO_2, SiO_2, SnO_2, and PbO_2. Therefore, since all the elements in a group have similar characteristics, he said element 32 would have an oxide formula of XO_2. By observing the other properties of its neighbors in the periodic table, he predicted several other properties of the unknown element (see Table 13-6).

Following the formulation of the periodic table, an intensive search for the missing elements was launched, and it was so successful that almost a dozen new elements were discovered within fifteen years' time. The table thus served as an excellent guide in the search for new elements. We do not use the table for this purpose anymore, however, because elements for all the empty spaces in the table have been found.

Another significant function of the periodic table is its use as a means of classifying and systematizing a tremendous amount of scientific information. Much of the descriptive information—facts and observations—about the elements can be correlated in the table. For example, the solubilities of the sulfate salts of the elements in Group IIA decrease in going down the group. The solubilities of these salts in 100 ml of water are:

Magnesium sulfate	26.9 g
Calcium sulfate	0.1619 g
Strontium sulfate	0.0113 g
Barium sulfate	0.0002 g

Another example: Since iron, cobalt, and nickel in Group VIII are highly colored in salts, we would expect the other elements in this group also to form highly colored salts. All

nine of the elements in Group VIII do, in fact, form complex compounds that are closely related to each other.

The periodic table has also been a valuable tool in the research into the use of metals. The usefulness of many metals is greatly increased when they are alloyed with other metals. With dozens of metals to choose from, however, it would be an almost endless task to try all the possible combinations in order to find the ideal ones. The table, by placing similar elements in the same group, suggests certain combinations of metals that are likely to give us what we want. For instance, the metal niobium (Nb) in Group VB improves the hardness and corrosion resistance of steel; the resulting alloy is ideal for surgical and dental instruments. But niobium is expensive, and, in order to reduce the cost, metallurgists turned to another metal in Group VB. By substituting tantalum (Ta) for niobium in the alloy, they produced a steel that is almost as good as the niobium steel. Cheap, useful catalysts have also been found by means of the periodic table. The uses of iron as a catalyst in manufacturing ammonia and of tetraethyl lead in gasoline stem from studies of the position of iron and lead in the periodic table.

Another common use of the periodic table is in classifying, or grouping, elements that have similar properties. The noble gases appear in the column on the far right. Next, in Group VIIA are the halogens—fluorine, chlorine, bromine, iodine, and astatine. These elements are so similar in many respects that they are called a *family*. The coinage metals, copper, silver, and gold, are grouped together in the same column. The alkali elements—lithium, sodium, potassium, rubidium, and cesium—are in Group IA, and the alkaline earth elements—magnesium, calcium, strontium, and barium—are in Group IIA; each of these groups of elements has many things in common in both physical as well as chemical properties. In two rows at the bottom of the main table are two series of elements for which there is no space in the table proper. Along with element 57, lanthanum, are thirteen other elements, all of which have very similar properties, and we might think of all of these elements as being packed into the space in Group IIIB of Period 6. Likewise, the actinide series of elements appear to be much alike, and they too may be viewed as fitting into a single space in the main table—the space in Group IIIB of Period 7.

In its broadest implication, the universal nature of the periodic table is evidenced by the composition of the sun and stars. From their spectra which serves as a key to their composition, all the elements identified in the universe are already included in the periodic table. The space traveler may find strange objects and unusual sights on his voyages, but the elements he finds in these strange places will very likely be identical to those already familiar to us through the table.

The periodic table, however, is by no means perfect. It has several defects and some inconsistencies that might lead

the unwary astray. For instance, it is difficult to place the short periods and long periods in the table so that all the interrelationships that are valid appear. There is no proper place for hydrogen, and the lanthanide and actinide elements are fenced off by themselves. The atomic weights of three pairs of elements are inverted—argon (18) and potassium (19), cobalt (27) and nickel (28), and tellurium (52) and iodine (53). Most of all, it is difficult to correlate the position of all the elements in the reactivity series (Table 13-1) with their positions in the periodic table. But in spite of these defects, the table must be given credit for establishing many of the basic ideas that were to become the foundation for our modern studies of atomic structure.

13-7 Metals and Nonmetals

As you have probably noticed, the periodic table is divided into two general groups—metals and nonmetals. We all know that metals such as copper, gold, iron, and chromium are much different in appearance from the nonmetals, sulfur and carbon (charcoal). But what is the scientific difference between them? Since there is no simple definition of either, the best way to differentiate them is to list their differences. Metals have luster, gloss, and brilliance; nonmetals are usually dull. Metals are good conductors of electricity; nonmetals are generally good insulators. We saw how this difference follows from the fact that metals have electrons with considerable freedom to move, whereas the electrons in nonmetals are tightly bound to their atoms. Metals are good heat conductors; nonmetals are poor heat conductors. Other distinctions between metals and nonmetals lie in their mechanical properties. Metals are ductile and malleable (capable of being drawn out into fine wires or fine sheets); nonmetals are not. Metals have tensile strength (the ability to resist elongation); solid nonmetals are brittle.

In addition to the physical differences between metals and nonmetals, there are important chemical differences to which we shall refer later. Metals and nonmetals, in fact, can be most easily classified by their chemical properties. The metals are located on the left side of the periodic table, the nonmetals on the right. Since there is no abrupt division between the two classifications, many elements have properties characteristic of both. To the far left are the light metals, the most reactive metals of all, which become still more reactive toward the bottom of the table. Thus, the most reactive metal is element 87, francium (Fr), with cesium (Cs) and rubidium (Rb) following closely behind.

The central part of the table contains the so-called heavy metals, whose chemical and physical properties vary with their positions in the table. As we go to the right, the elements become less metallic and gradually give way to the nonmetals. The elements near the borderline have intermedi-

ate properties. For instance, aluminum, commonly thought of as a metal, shows definite nonmetallic characteristics at times. At room temperature, aluminum is ductile and malleable, a typical metal. But at temperatures above 200°C, it becomes brittle enough to be ground into a fine powder that serves as the pigment in aluminum paint. As we have just seen, brittleness is a property of nonmetals. In chemical reactions, aluminum also often acts like a nonmetal.

To the far right of the table are the nonmetals. Since nonmetallic properties are most noticeable among the elements in the upper right corner of the table, the most nonmetallic element is fluorine (F). Note the interesting gradations in Group VA. The elements at the top are entirely nonmetallic —nitrogen (N) and phosphorus (P) have no metallic properties. Arsenic (As), however, has some metallic characteristics, antimony (Sb) more, and bismuth (Bi) still more. Thus, within one group of the table, the elements range from a true nonmetal to a fairly typical metal.

13-8 The Periodic Table and Atomic Structure

In the last chapter we saw how modern science pictures the atom as a nucleus surrounded by energy levels containing 2, 8, 18, 32, and so on electrons. The number of these levels in any particular atom depends upon the atomic number—that is, the number of electrons in the atom. Although the periodic table preceded the present concept of atomic structure, it has greatly aided the scientist in his investigation of the electronic arrangement of the atom.

In Fig. 13-6 the first 20 elements are shown, arranged in the order of their increasing atomic numbers and in rows and columns that are identical to those in the periodic table. Around each element is a circle on which is represented the electrons in the outermost, stable energy level of its atom. If we compare the number of electrons in this energy level with the group number for the element, we find that they are identical. Thus, in Group I all the elements have one

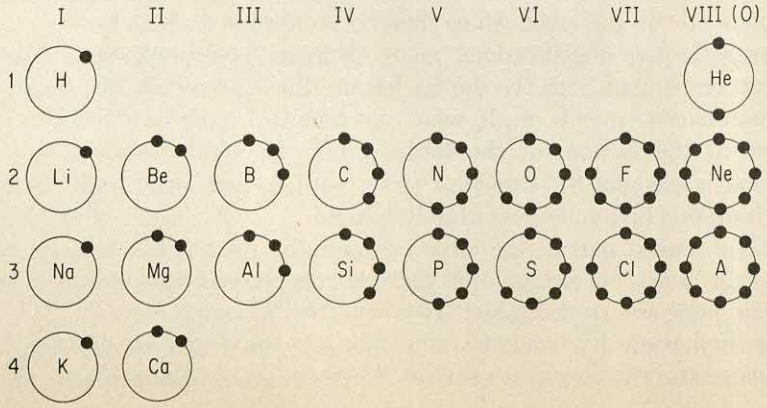


Fig. 13-6. The outer electronic structures of the lighter elements and their positions in the periodic table.

electron in their outer energy levels. We can now see that the elements in any given group are similar largely because their electronic structures are so much alike. The fact that Group II has two electrons in the outer energy level, Group III three electrons, and so on provides additional proof that our picture of the atom is a dependable one. We also have the reassuring knowledge that the information gained through investigations of electronic structure corroborates the arrangement of the elements in the periodic table.

The first period in the table (see Fig. 13-4), containing hydrogen and helium, corresponds to the first energy level in atomic structure. The second period contains 8 elements, lithium through neon, and each of them corresponds to a successive addition of one electron to the second energy level. The third period containing 8 elements, sodium through argon, is similar to the second period. In the fourth period there are 18 elements, 8 of which are accounted for by the addition of 8 electrons to the fourth energy level. But once there are electrons occupying the fourth energy level, an additional 10 electrons can be accommodated in the third level. Again, in the fifth period there are 18 elements—8 electrons have been added to the fifth energy level and 10 electrons have been added to the fourth energy level. The sixth period contains 32 elements. In terms of atomic structure, these are accounted for as follows: 8 electrons are added to the sixth, 10 to the fifth, and 14 to the fourth energy levels. We see here the operation of the rule that no energy level can contain more than 8 electrons when that level is the outermost occupied level. But the maximum number of electrons in the third level goes up to 18 electrons when there is one more level occupied, and up to 32 electrons in the fourth level when there are two more levels occupied. Thus, the short and long periods in the table are explained by these irregularities in atomic structure. Since there are seven major energy levels in the structure of the atom, there are seven periods in the periodic table.

Even though the relationship between electronic structure and the periodic arrangement grows more complex beyond the third period, these irregularities are not insurmountable. In fact, the table has served well in guiding the study of the atom's structure, and as the table becomes more complex it merely reflects the more complex structures of atoms. The parallel nature of the table and atomic structure shows that there is a basic relationship between them—one complements the other.

13-9 Summary

In general, the elements that are the easiest to obtain from their ores are the least reactive and have been known for the longest time. The most reactive elements require the most vigorous methods to remove them from their ores. The

reactivity series of the elements provides a means of classifying elements upon the basis of the tendency they have to enter into chemical changes.

The most abundant elements in nature are those with relatively low atomic numbers and those with an even-even neutron-proton count in their nuclei. The rarest elements are those having odd numbers of neutrons or protons.

The periodic classification of the elements is based on an arrangement of the elements in tabular form, with the elements listed in the order of increasing atomic numbers. In such a listing, there is a periodic recurrence of similar properties, which is the basis for the periodic law. The elements in horizontal rows are called periods, and those in vertical columns are called groups. With the elements organized into these periods and groups, we can correlate many of their chemical and physical properties, as well as their atomic structures. The advantages of the table far outweigh its defects, and the table still occupies a foremost position among the tools of the modern scientist.

EXERCISES

A. KEY TERMS TO REMEMBER

alkali
alkaline earth
family of elements
group
halogen
lithosphere
metal
nonmetal
period
periodic classification of elements
periodic law
reactivity series

B. QUESTIONS ABOUT THE ELEMENTS

1. Give the derivations of the names of the following elements: lithium, rubidium, rhodium, cesium, ytterbium, vanadium, tantalum, palladium, iridium, and indium. Check your guesses in a good dictionary.

2. Under what conditions is the process of obtaining metals from sea water economical? Name several methods by which various substances can be removed from sea water.

3. Describe three geological processes by which ores are collected into useful deposits.

4. Which elements are more abundant in the earth's crust, those with an even or an odd number of protons in their nuclei? Can you give a reason for this?

5. At what distance does the law of coulombic force fall down?

6. State the periodic law and give an example of its application.

7. From the periodic table, find the following information:
   a. the element most nearly like iridium (Ir).
   b. the element most nearly like carbon (C).
   c. the formula for magnesium sulfate, given that the formula of calcium sulfate is $CaSO_4$.

d. the hydrogen-sulfur compound similar to water (H_2O).
e. three elements that are very unstable and that decay into radioactive products.

8. List three uses of the periodic table.

9. Compare a period in the periodic table to an energy level in atomic structure, and show why the periodic table takes its form from the electronic structure of the elements.

10. Suppose that an astronaut returned to earth with an element whose atoms have a negatively charged nucleus and positively charged outer part. If its mass is 40 and its atomic number is 20, sketch a structure for it. Predict how this element would behave if it came into contact with one of our earth elements?

11. From a chemical handbook, or other reference, find the (a) melting points, (b) boiling points, and (c) densities (or specific gravity) of the first 36 elements. Plot these data similar to Fig. 13-5 in order to test the principle of periodicity of these properties.

Fourteen

Chemical Bonds

In the last chapter we noted two concepts that are fundamental to the elements—the reactivity series and the periodic law of the elements. Both of these concepts are directly related to the actual atomic structure of the elements, and, as a consequence, in order to understand why the elements show wide variations in reactivity and also fall into an orderly periodic arrangement, we must re-examine the atoms themselves.

Early in the study of chemical reactions, two very important laws were derived from laboratory experiments: the law of conservation of mass and the law of definite proportions (see p. 206). As you will recall, the second of these two laws states that in any given compound, regardless of its source, the elements are always united with each other in exactly the same weight ratio. In other words, a particular compound always has exactly the same composition. Scientists, in pursuing their questioning behind these laws, found that the concept of the electronic structure of the elements explains how both these laws operate. It is interesting to note that these laws, and much more about the chemical composition of substances was worked out long before the modern concepts of atomic structure were established. Certain elements, for example, such as sodium, have been known for a long time never to occur free in nature. Now we are in a position to understand why sodium readily forms chemical bonds with other elements, and in this chapter we will examine certain factors of atomic structure that explain why sodium is a very reactive element.

TABLE 14-1 ATOMIC RADII OF THIRD PERIOD ELEMENTS

Element	*Atomic Radius*	*Number of Outer Electrons*
Na (sodium)	1.86 A	1
Mg (magnesium)	1.60	2
Al (aluminum)	1.48	3
Si (silicon)	1.17	4
P (phosphorus)	1.0	5
S (sulfur)	1.06	6
Cl (chlorine)	0.97	7

14-1 Atomic Radii

We saw in Fig. 13-5 that there is a periodic change in the size of the atoms. Those elements in Group IA of the periodic table have by far the largest atoms in any horizontal row, and as we move to heavier elements in any row, the atoms become smaller. For example, in the third period, starting with sodium, the elements decrease in atomic radii as shown in Table 14-1. Similar relationships occur in other periods, although minor variations appear in the longer periods which need not concern us here.

Why should the atoms decrease in size as they become more complex? We must recall that in building up the atoms, each increase in atomic number represents additional electrons in the outer part of the atom and additional protons in the nucleus. There will also be additional neutrons to account fully for changes in atomic weights. In the third period, for example, sodium has 11 electrons, 11 protons, magnesium has 12 electrons, 12 protons, and so on. The electrons are being added to the third principal energy level and the protons are being added to the nucleus. This increased nuclear charge exerts a greater attraction on the electrons with the result that the electron clouds shrink in size. Thus, we may expect to find chlorine's nucleus exerting a much stronger attraction on its electrons in the third energy level than the sodium's nucleus exerts on its electron in the same energy level. Incidentally, as the third energy level shrinks, there will be a proportionate shrinkage of the inner energy levels.

When we examine elements within a group or vertical row of the periodic table, the atomic sizes increase as more energy levels are occupied. Even though there are greater nuclear charges within the heavier elements, the higher energy levels are sufficiently larger to account for this effect. Table 14-2 summarizes these changes for the Group I elements.

14-2 Valence and Electrolysis

One of the fundamental properties of the elements is their ability to gain or lose electrons, or to share electrons with

TABLE 14-2 ATOMIC RADII OF GROUP I ELEMENTS

Element	*Atomic Radius*	*Number of Outer Electrons*
Li (lithium)	1.50 A	1
Na (sodium)	1.86	1
K (potassium)	2.27	1
Rb (rubidium)	2.43	1
Cs (cesium)	2.62	1

each other in order to achieve a greater amount of stability. Long before there was an understanding of the electronic structure of atoms, however, a measure was derived of the capacity elements have to react with one another. This capacity is traditionally called *valence,* and it was proposed by Edward Frankland of England over a century ago. Valence is expressed by a small, whole number and it represents the number of electrons an atom gains, loses, or otherwise contributes (shares) to form a chemical bond. Positive valence numbers signify the loss of electrons and negative numbers signify the gain of electrons, and Table 14-3 gives

TABLE 14-3 VALENCES OF SOME COMMON ELEMENTS

Sodium	Na	+1
Potassium	K	+1
Hydrogen	H	+1
Magnesium	Mg	+2
Calcium	Ca	+2
Zinc	Zn	+2
Aluminum	Al	+3
Carbon	C	+4, −4
Silicon	Si	+4, −4
Nitrogen	N	+5, −3
Oxygen	O	−2
Sulfur	S	−2, +6
Chlorine	Cl	−1, +7
Copper	Cu	+1, +2
Silver	Ag	+1

the valences of several common elements. By using the following simple rule, we can employ these valance numbers to write formulas of compounds: The valence numbers of the elements in a compound must add up to zero, that is, there should be an equal amount of positive and negative valence in a compound. Let us examine a few examples. We have already encountered the compound, sodium chloride, and its formula, NaCl, and if we write the valences for the sodium and chlorine above the formula, as follows,

$$\begin{matrix} +1 & -1 \\ \text{Na} & \text{Cl} \end{matrix}$$

we see that the sum of these valences (+1 and −1) is zero.

In sodium oxide, Na_2O, we have:

$$\underset{Na_2}{2(+1)} \qquad \underset{O}{-2}$$

Since in this compound there are two sodium atoms, the total valence of sodium is twice the valence of each atom. The sum of the valences in sodium oxide ($+2$ and -2), therefore, is zero.

In the compound, sulfuric acid, H_2SO_4, the valence numbers are:

$$\underset{H_2}{2(+1)} \quad \underset{S}{+6} \quad \underset{O_4}{4(-2)}$$

and in hydrogen sulfide they are:

$$\underset{H_2}{2(+1)} \qquad \underset{S}{-2}$$

The disparity in the valence of sulfur in the two compounds illustrates the fact that some elements have more than one valence number. Note that in each compound the total valence of the molecule is still zero. Although certain elements are known to have four or five different valence numbers, an element usually has only one or two.

The basis on which valences are derived stems from a variety of experimental data dealing with chemical reactions. Some of the most useful data for this purpose are found in reactions performed by electrolysis (from the Greek words *electro* and *lysis*, meaning to loosen or decompose by electricity). Electrolysis reactions have long been a favorite type of chemical change, both for the scientists interested in fundamentals and the practitioner seeking methods of making useful products. For example, the production of copper, magnesium, aluminum, chlorine and many other elements involves electrolysis on a large scale, and the plating of many metals such as chromium, silver, and gold is customarily achieved by electrolysis.

One of the earliest scientists to investigate electrolysis was Michael Faraday in England, who, during the 1830's, advanced some very useful concepts that are known as *Faraday's laws*. These concepts may be summarized in the statement: the quantity of electricity in coulombs (q) equals the number of equivalents (n) of an element that undergoes a change times the number of faradays (F), and this statement is expressed in Equation 14.1:

$$q = nF \tag{14.1}$$

A faraday is 96,500 coulombs or 6.02×10^{23} electrons. For instance, Faraday found that one ampere of current in ten seconds plated out exactly twice as much silver as did one ampere of current in five seconds. Further, 27 amperes in one hour liberates 9 g aluminum, 12 g of magnesium, or 23 g of sodium (Fig. 14-1). The quantity of sodium released in Fig. 14-1, 23 g, is one gram-atomic weight, which we know to con-

tain Avogadro's number of atoms— 6.02×10^{23} atoms. The amount of current that flows in an hour at 27 amp $\times$ 60 $\times$ 60, or approximately 96,500 coulombs. Since a coulomb is the flow of 6.23×10^{18} electrons (see p. 161), 96,500 coulombs is the flow of 6.23×10^{18} electrons $\times$ 96,500 coulombs, or 6.02×10^{23} electrons. Thus, 6.02×10^{23} atoms of sodium are liberated by 6.02×10^{23} electrons (96,500 coulombs), and we may thus conclude that each sodium atom requires one electron to release it from its compound, sodium chloride. In this process, we obviously have an extremely accurate method for counting atoms.

But with magnesium and aluminum, the results are a bit different. Twelve grams of magnesium make up only half a gram-atomic weight. Because 12 g of magnesium contain only ½ of 6.02×10^{23} atoms, each magnesium atom requires *two* electrons to liberate it from its compound. And since only 9 g of aluminum (one-third the gram-atomic weight of aluminum) are liberated by 6.02×10^{23} electrons, each aluminum atom requires *three* electrons to liberate it from its compound.

From the results of the experiment in Fig. 14-1, we can define the term *equivalent weight,* used in Faraday's law, as the weight of an element that will be liberated by 96,500 coulombs of electrical current. With this definition, and the valences of sodium, magnesium, and aluminum, we can derive a simple relationship: The equivalent weight of an element is its atomic weight divided by its valence number.

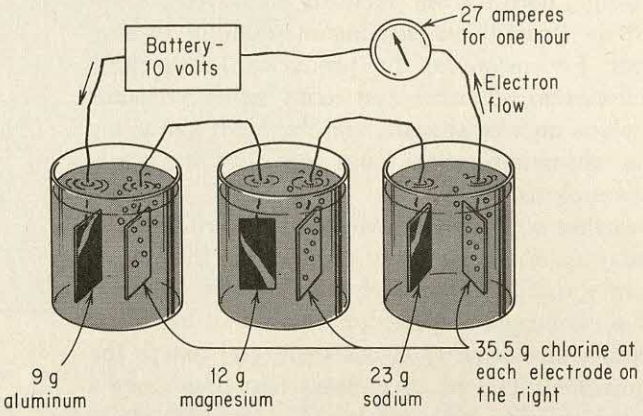


Fig. 14-1. The quantities of aluminum, magnesium, and sodium liberated by the electrolysis of their fused salts.

Faraday's work definitely confirmed the validity of Avogadro's number. If we know the valence of an element and count the number of electrons that are passed through an electrolysis cell, we can determine the weight of the element, and the number of atoms, plated out by weighing the electrode before and after the experiment. The number of silver atoms (the valence of silver is +1) so released, for example, will be exactly equal to the number of electrons passed through the electrolysis cell. Such an experiment provides further proof of the accuracy of Avogadro's number.

Faraday's method also gives us the valence of an element when that element has more than one valence. For instance, copper has two valences, +1 and +2. If we wish to plate out, say, 10 g of copper from a copper chloride solution, the amount of current required depends upon which valence of copper is present. In cuprous chloride, CuCl, copper's valence is +1, but in cupric chloride, $CuCl_2$, copper's valence is +2. To plate out the 10 g of copper from $CuCl_2$ requires twice the number of electrons that is required to plate out 10 g of copper from CuCl. This shows experimentally that copper has two different valences because a different amount of electrical current is required to plate out equal amounts of copper from these two copper compounds. There is an obvious relationship between the valence of an element and the number of electrons required to plate it out: In elements having a valence of +1, each atom requires one electron; in those with a valence of +2, each atom requires two electrons; and for a +3 valence, each atom requires three electrons.

If we were to examine those elements with a negative valence, we would find a similar relationship. Each atom of an element with a valence of −1 *gives up* one electron when it plates out of solution, and the atom of an element with a valence of −2 gives up two electrons. Each atom of oxygen plated out by electrolysis, for example, gives up two electrons to the positive electrode as it is released from the water in which it was originally combined.

STATE INSTITUTE OF EDUCATION Govt. of West Bengal

14-3 Ionization Potential and Electron Affinity

In order to understand why the elements exhibit various valences and what kind of chemical bonds they form, we must recall the electronic structures of the elements, their atomic radii, and two new ideas—the ionization potential and the electron affinity of the elements. For those elements that tend to lose electrons when they react chemically, such as sodium, calcium, and iron, we need to know a measure of the work required to remove completely an electron from one of their atoms. This amount of work, already described on page 231 is called the *ionization potential.* Thus, to remove the outer electron from a sodium atom (measured in the gaseous state) requires 5.14 electron volts, and the ionization potentials of the elements in Group I of the periodic table are given in Table 14-4. Thus, it is apparent that the larger the atom the less work is required to remove the valence electron from it. Other factors, such as the nuclear charge and the shielding effect of other electrons in the atom influence the ionization potential, but these need not be of particular concern to us here.

The electron affinity of an element is the amount of energy released or absorbed when an electron is added to an atom. For the chlorine atom in the reaction $Cl + e^- \rightarrow Cl^-$, the

TABLE 14-4 ATOMIC RADII AND IONIZATION POTENTIALS OF GROUP I ELEMENTS

Element	*Atomic radius*	*Ionization potential*
Li	1.50 A	5.39 ev
Na	1.86	5.14
K	2.27	4.34
Rb	2.43	4.18
Cs	2.62	3.89

TABLE 14-5 ELECTRON AFFINITIES OF GROUP VII

Element	*Atomic radius*	*Electron affinity*
F (fluorine)	0.72 A	3.63 ev
Cl (chlorine)	0.99	3.78
Br (bromine)	1.14	3.54
I (iodine)	1.33	3.24

electron affinity is 3.78 electron volts (Table 14-5), and this process may be regarded as the opposite of the process involved in measuring the ionization potential. However, there is an inherent difference in the elements on the left-hand part of the periodic table (the metals) and those on the right-hand part (the nonmetals). Metals tend to lose electrons when they react chemically, and nonmetals tend to gain electrons. Thus, the valence numbers of the metals in Table 14-3 are positive for the metals (electron losers) and negative for the nonmetals (electron gainers).

The ionization potentials and the electron affinities are measures of the tendencies of elements to react chemically. It is safe to presume that a metal with a low ionization potential and a nonmetal with a high electron affinity will react vigorously with each other and form a stable product. For example, from Tables 14-4 and 14-5 we may predict that cesium fluoride is more stable than lithium iodide because of the greater tendency for electron transfer in CsF than in LiI. There are other factors that would influence such a prediction, and in order to accommodate all the factors known about the elements and their tendencies to form chemical bonds, a convenient system called *electronegativities* (Fig. 14-2) has been devised. This is an arbitrary scale, but it de-

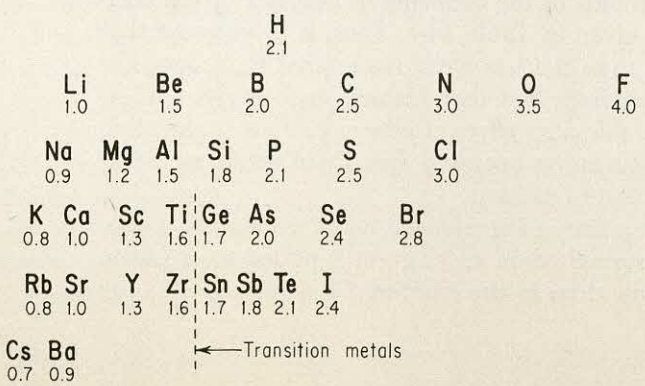


Fig. 14-2. Electronegativities of the elements.

pendably serves as a guide to the kind of chemical bonds two elements will form. We will make use of this guide as we now turn to a discussion of the kinds of chemical bonds encountered in chemical compounds.

14-4 Electrovalence

An important principle based upon the best we know about atomic structure states that the electrons in the atom always seek their lowest energy level. We encountered this principle when we studied the electrons in excited atoms—the excited electrons always return to their lowest energy level when the source of excitement is removed. Furthermore, among the lighter elements the innermost energy levels are filled with electrons before the outer levels. When applied to groups of atoms, the principle says that if two atoms unite with each other they do so in order to permit their electrons to attain a lower energy level. When they unite, the atoms either lose or gain electrons in an attempt to fill their outer energy levels and attain a more stable electronic arrangement. The simplest means of accomplishing this goal is the complete transfer of an electron from one atom to another—a process that is called *electrovalence.* Since those elements whose outer level is already filled—the inert gases such as helium, neon, and argon—need not make any electron transfers to reach stability, they never form any compounds. But most of the elements whose atoms have less than eight electrons in their outer energy levels seek a more stable electronic arrangement by combining with other elements.

The sodium atom has one electron in its outermost energy level (Fig. 14-3), and it gives up this electron quite easily. When the sodium atom loses this electron, the atom becomes a positively charged ion. An *ion* is an atom, or a group of atoms, that has either lost or gained electrons. Since the chlorine atom has seven electrons in its outermost energy level, the most direct means of attaining electronic stability is the gain of a single electron. When a chlorine atom takes on an electron, the resulting chloride ion has eight electrons in its outer energy level. This arrangement, you will note, is the electronic structure of the argon atom. However, the chloride ion is not identical with the argon atom in all respects—only the electron cloud of the chlorine is affected, its nucleus remains unchanged in this process.

If an atom that tends to lose an electron meets an atom that tends to gain an electron, such as sodium and chlorine in Fig. 14-3, the electron will immediately be transferred to enable the atoms involved to become more stable. The reaction between sodium and chlorine is primarily a transfer of electrons from the sodium to the chlorine, and the resulting compound, sodium chloride, consists of an equal number of sodium ions and chloride ions. The presence of tremen-

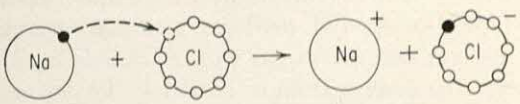


Fig. 14-3. The sodium atom loses its outermost electron, while a chlorine atom gains an electron in order to fill its outer energy level.

dous quantities of sodium chloride in the ocean and in the numerous salt deposits in the earth's crust attests to the stability that is achieved through the chemical union of sodium atoms with chlorine atoms.

Magnesium also reacts with chlorine through an electron transfer. The atomic structure of magnesium shows that there are two electrons in its outer energy level (Fig. 14-4). Since chlorine requires only one electron to reach stability, when magnesium and chlorine react, one magnesium atom furnishes the electrons needed by two chlorine atoms. Magnesium chloride, therefore, contains one magnesium ion for each two chloride ions, and its formula is $MgCl_2$. In aluminum chloride, each aluminum atom supplies three electrons (Fig. 14-4), one for each of the three chlorine atoms.

The strength of the electrovalent bond between two atoms or ions can be gauged by their difference in electronegativity (Fig. 14-2). The greater the difference in the electronegativities of two elements, the stronger is the bond between them. For example, this difference for sodium and chlorine is 3.0 — 0.9 or 2.1, and for magnesium and chlorine is 3.0 — 1.2 or 1.8. Thus, we may conclude that the chemical bonds in NaCl are somewhat stronger than those in $MgCl_2$. When the electronegativity difference between two elements is smaller than 1.7 (it is 1.5 for $AlCl_3$) the chemical bond takes on another aspect, or it is said to be less electrovalent and more covalent in its character. We now turn to the covalent chemical bond.

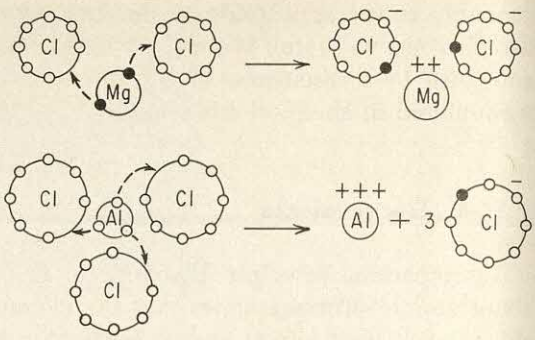


Fig. 14-4. The formation of magnesium chloride and aluminum chloride through electron transfer.

14-5 Covalence

Covalence is the sharing of electrons by atoms in order to achieve a stable electronic structure. This kind of valence is much more common than electrovalence, although it does not usually lead to compounds that are as stable as those formed by electrovalence.

Because the carbon atom, atomic number 6, has four electrons in its valence level, it must either gain four electrons of lose four electrons to achieve electronic stability. It is such a rare event for an atom to gain or lose this many electrons that carbon must find electronic stability by *sharing* its electrons with other atoms. If we examine the structure of the carbon compound, methane, a primary compound of natural fuel gas and marsh gas with a formula of CH_4, we will find that the carbon atom is located at the center of the molecule and is attached to the four hydrogen atoms by covalent bonds (Fig. 14-5). Stability is thereby achieved because the carbon atom now has eight electrons in its valence level—its four original electrons plus one from each of the four hydrogen atoms—and each hydrogen atom has fulfilled its tendency to complete its outer valence level of two electrons.

The simplest covalent molecule is hydrogen, H_2. Under ordinary conditions, hydrogen occurs as a diatomic molecule

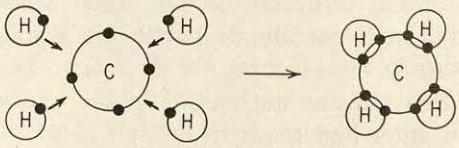


Fig. 14-5. The electronic structure of methane, CH_4.

in order to achieve a maximum amount of stability (Fig. 14-6), and the relationship between the two atoms can be explained in terms of Coulomb's law of electrical attraction and repulsion. Since each hydrogen atom, as we know, consists of a single proton and a single electron, when two such atoms approach each other, four different forces are produced (Fig. 14-7): A force of repulsion exists between the two nuclei and between the two electrons, and a force of attraction is set up between the electron of one atom and the nucleus of the other atom. The relative sizes of these forces are shown in the graph in Fig. 14-7. At a certain distance, the force of attraction between the atoms is at a maximum, or, to put it another way, the potential energy of the two atoms is at a minimum. For the hydrogen molecule, this distance of maximum stability is 0.74 A. The two electrons are thus held jointly by the two nuclei, and this sharing constitutes the chemical bond that holds the two hydrogen atoms together—indeed, this is the bond that holds all covalent molecules together.

The difference between electrovalence and covalence is mostly a matter of degree, particularly when the electronegativity difference between the two elements is an intermediate value. Purely covalent bonds involving equal sharing of the electron pair between two atoms occurs only when identical atoms are involved, as in H_2, Cl_2, N_2, etc. But with the strongest electrovalent bonds the electron transfer is not 100 per cent complete. That is, in NaF, where the electronegativity difference is 3.1, the bond is about 82 per cent electrovalent and 18 per cent covalent in character.

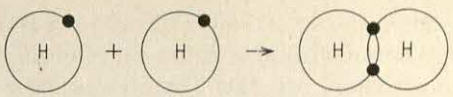


Fig. 14-6. The hydrogen molecule.

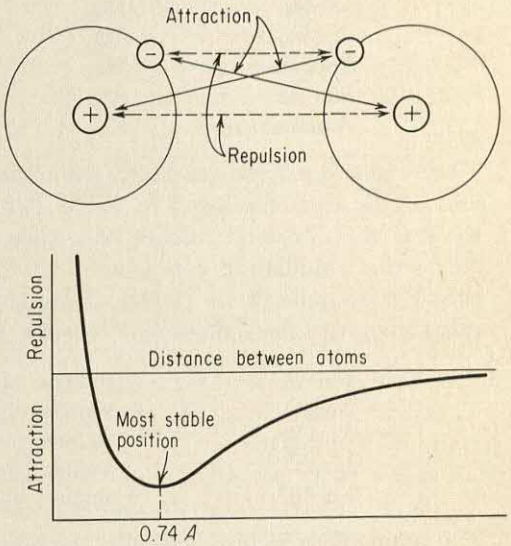


Fig. 14-7. The forces acting between two hydrogen atoms.

14-6 Coordinate Valence

As we become acquainted with the formulas of more complex substances, we encounter a third kind of valence, in addition to electrovalence and covalence. Many of these substances contain within their molecules distinctive groups of atoms that often remain intact, even though a vigorous chemical change takes place. These groups of atoms, called *radicals*, are often bonded together with a special kind of covalence, called *coordinate valence*. In ordinary covalence, each atom involved contributes an electron to the pair that is shared between the two atoms, but in coordinate valence, both electrons are contributed by one of the atoms, while the other atom comes into the bond and shares the electrons donated by the other. Both atoms achieve a more stable electronic arrangement than they had prior to their bonding.

An example of a substance that contains a radical is sulfuric acid, H_2SO_4. When this acid is added to a solution of calcium chloride, $CaCl_2$, a white insoluble substance, calcium sulfate, is formed:

$$H_2SO_4 + CaCl_2 \rightarrow 2HCl + CaSO_4$$

Note that the sulfate radical (SO_4) goes through the chemical reaction unchanged. From the valence of either hydrogen or calcium, we find that the valence of the sulfate radical is -2. A *radical,* therefore, is an ion that takes part in a chemical change as if it were a single atom; it exhibits a regular valence and behaves as a chemical unit.

Here are several common radicals along with their valences:

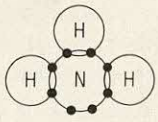
Sulfate	SO_4	-2
Nitrate	NO_3	-1
Carbonate	CO_3	-2
Phosphate	PO_4	-3
Silicate	SiO_3	-2
Ammonium	NH_4	$+1$

These radicals may be combined with other ions, such as the ions of the elements listed in Table 14-3. Bearing in mind the rule that the total valence of a molecule is neutral, we may write formulas of compounds that contain these radicals. For example, if we combine each of the negative radicals above with the sodium ion (valence $+1$), we have:

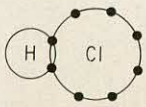
$NaNO_3$	sodium nitrate
Na_2SO_4	sodium sulfate
Na_2CO_3	sodium carbonate
Na_3PO_4	sodium phosphate
Na_2SiO_3	sodium silicate

The ammonium radical logically combines with a negative ion, such as chloride (valence -1), to give ammonium chloride, NH_4Cl.

Radicals are usually bonded to the rest of the molecule by electrovalence, but the atoms within a radical are more often held together by a combination of covalence and coordinate valence. To illustrate this concept, suppose we examine an ammonia molecule, represented by the formula, NH_3, the structure of which is shown in the diagram:



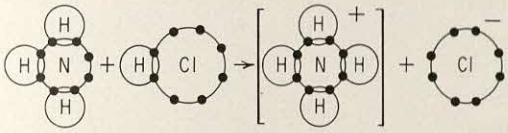
The nitrogen atom shares a total of three electrons with the three hydrogen atoms and thus completes its valence level. Each hydrogen atom, with two electrons, is stabilized by its covalent bond with the nitrogen. But, note that a pair of electrons in the outer energy level of the nitrogen atom is not bound in any interatomic bond and is therefore available to form a coordinate valence bond with another atom. If we add to the ammonia molecule a molecule of hydrogen chloride, HCl, which is a covalent molecule with the structure:



we obtain the product ammonium chloride.

$$NH_3 + HCl \rightarrow NH_4Cl$$

Since ammonium chloride is an electrovalent compound, composed of the ammonium radical and the chloride ion, let us see what happens when this reaction takes place. The ammonia molecule, NH_3, has now become the ammonium radical NH_4^+. An electronic diagram will help us explain this change. When the hydrogen moves from the hydrogen chloride to the ammonia molecule, it leaves behind the two

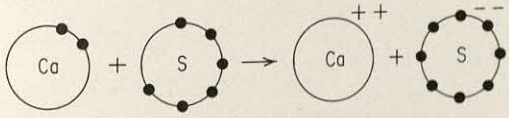


electrons it shared with the chlorine because the chlorine atom has a greater attraction for these electrons than does the hydrogen atom. The stripped hydrogen atom is only a proton, so it readily forms a bond—called a *coordinate bond*—with the free pair of electrons on the nitrogen atom and becomes stabilized in the ammonium radical. Once the coordinate bond has formed, it appears no different from the other bonds in the ammonia molecule. The addition of a proton to an otherwise neutral molecule imparts a positive charge to it and makes the net charge, or valence, of the ammonium radical a positive one.

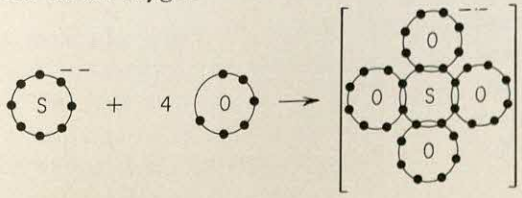
The sulfate radical, another example of coordination valence, is composed of sulfur and oxygen, but the sulfur atom must first be changed into an ion with its valence level filled. This change may be accomplished by igniting a mixture of calcium and sulfur to produce a vigorous reaction from which a compound called calcium sulfide is obtained. This reaction may be shown by the equation:

$$Ca + S \rightarrow CaS$$

Two electrons are transferred from the calcium atom to the sulfur atom:



Calcium sulfide is an electrovalent compound in which both calcium and sulfur have an electronic structure like that of argon, element number 18. That is, the calcium has lost two of its original 20 electrons, and sulfur has added two electrons to its original 16; of course, nothing has happened to the nuclei of these two atoms. The sulfur ion possesses four pairs of electrons in its outer energy level, and each of these pairs then enters into coordination valence with the six outer electrons of an oxygen atom to form a sulfate radical:

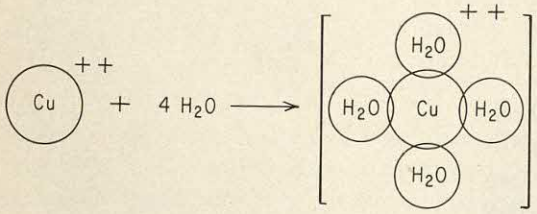


This radical, with a −2 electrovalence, is composed of five atoms held together by four coordination valence bonds; the central sulfur atom shares all its electrons with other atoms. Since the six electrons in the original sulfur atom are now shared through coordination valence, we can account for the valence of +6 for sulfur in Table 14-3. We know that the sulfate radical is a stable one because sulfates abound in the earth's crust; calcium sulfate, for instance, constitutes the minerals gypsum and anhydrite, two common sedimentary rocks.

The stability of other radicals, usually bound together by coordination valence, may also be gauged by the abundance of these radicals in a natural state. Tremendous quantities of the carbonate radical, combined with calcium, occur as calcium carbonate in limestone, shale, and slate. The most common occurrence of the phosphate radical is in calcium phosphate, sometimes called rock phosphate. Calcium phosphate forms a large part of our bones and teeth; a man of average weight has almost five pounds of calcium phosphate in his body.

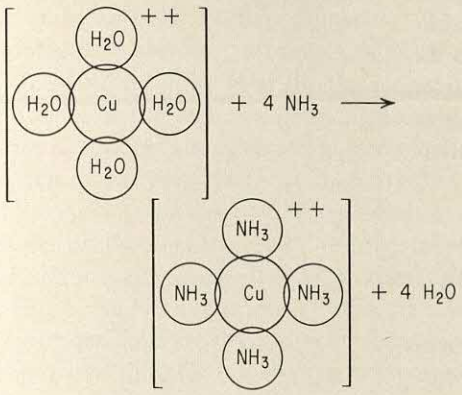
With the above examples in mind, we can now see that coordinate valence is an important type of chemical bond that holds many different kinds of molecules together. The central or coordinating atom may share one, two, three, or more pairs of electrons, and this number of pairs of electrons, or coordinate bonds, is called the *coordination number* of the central atom. Coordination numbers of four and six are the most common.

Coordination valence involving the metals produces some unique properties in these metals, properties that are caused by the fact that coordination bonds are present. Copper, a yellowish metal, becomes colorless when it forms the copper ion, Cu^{++}, and when this ion is added to water, the water is transformed into a beautiful azure blue. This blue color comes from the ion, $Cu(H_2O)_4^{++}$, a coordination system:



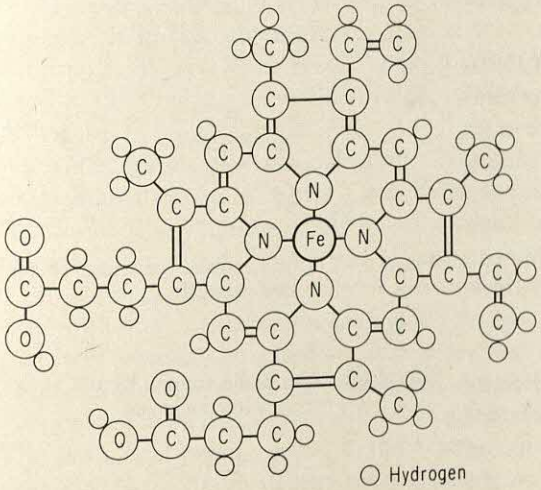
Since the water molecule has two unshared electron pairs in it (see Fig. 14-8), there are ample electron pairs from the four water molecules to form coordinate bonds with the central copper ion. It is believed that the bond between water and the copper ion is also the result of the strong dipolar nature of water attracted to the positively charged metal ion. This type of bond, called an ion-dipole bond will be more fully described in the next section. If we add ammonia to the copper-water complex, the pale blue turns to an intense, deep blue that is a distinctive characteristic of this

ion, which is frequently called the tetramminecupric ion. In this reaction, the four water molecules are merely displaced by a similar number of ammonia molecules:



The lone pair of electrons on the ammonia molecule as shown in the diagram on page 272 supplies the electrons used to make the coordination bonds in $Cu(NH_3)_4^{++}$.

Many dyes, pigments, and other colored materials are made up of coordination systems. The green color of all vegetation is due to chlorophyll, a complex molecule that has the magnesium atom at its center. The coordination number of magnesium in this colorful substance is 4—that is, the magnesium atom is surrounded by four nitrogen atoms. The red coloring matter in blood contains a substance similar to chlorophyll, except that it has iron as the central element instead of magnesium. This red substance, hemoglobin, can be separated into a protein called globin, and an iron coordination complex, hemin, shown below. In the



hemin molecule, which is related to chlorophyll in general structure, the central iron atom is attached to four nitrogen atoms. The four nitrogen atoms are, in turn, parts of five-membered rings to which are attached various radicals. Except for the central iron atom, the molecule is held together

with ordinary covalent bonds. In addition to the four nitrogen atoms, the central atom can be linked by coordinate bonds to the globin protein (structure unknown) and an oxygen or carbon dioxide molecule. It is the primary purpose of this complex substance to carry oxygen and carbon dioxide in biological processes, and the ability of hemoglobin to do this is based upon the coordination bonds of the central iron atom. The study of these complex, naturally occurring materials is one of the most exciting areas of present-day scientific investigation.

14-7 Other Types of Chemical Bonds

The world would be a dull and lifeless place if chemical bonds were restricted to electrovalence, covalence, and coordinate valence. Water would be a gas even at polar temperatures, and plant and animal life as we know it would be wholly impossible. Many of the common substances such as wood, paper, protein, and metals would be nonexistent because they are held together by bonds other than the three above. If we recall the principles of electrostatics in Chapter 8, we should have little difficulty understanding at least the rudiments of these additional but important bonds.

All of these additional bonds involve dipoles of molecular or atomic magnitude, and a dipole is simply a particle that has an imbalance of positive and negative charges. Thus, two dipoles attract each other by an electrostatic force which, according to Coulomb's law, attain considerable magnitude at the small distances between molecules, and this accounts for what is commonly called a dipole-dipole bond. A common substance whose properties are prescribed by dipole-dipole bonds is ordinary water and ice.

Suppose we examine a single molecule of water, H_2O, in order to ascertain its dipolar character. Oxygen, its central element, with an atomic number of 8, has six electrons in its outer energy level (Fig. 14-8). Each hydrogen atom shares its single electron with the oxygen atom, making the oxygen atom similar to the electronic structure of neon, and each hydrogen atom in turn has two electrons, making it similar to helium. The hydrogen and oxygen nuclei jointly hold those pairs of electrons that are between them.

In Figs. 14-8 and 14-9 note that the atoms in the water molecule do not lie in a straight line but form a definite angle of 105° (Fig. 14-9). Note also that the water molecule is not symmetrical, since it has more unshared electrons on the side away from the hydrogen atoms than it does on the side toward them. That is, if we examine the distribution of negative and positive charges in the water molecule, we will find that the centers of population for these two kinds of charges are not identical. The positive center lies on the side toward the hydrogen atoms, and the negative center lies quite near the center of the oxygen atom. The resulting im-

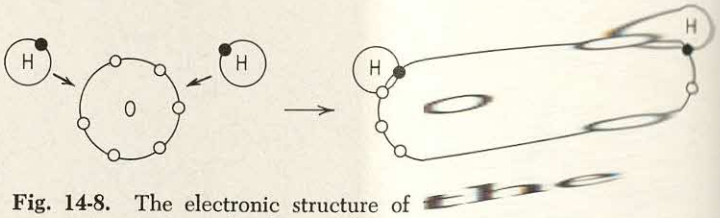
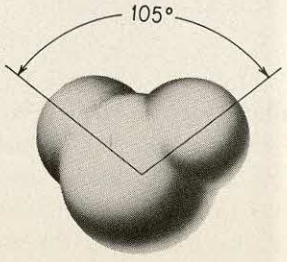


Fig. 14-8. The electronic structure of water molecule.

Fig. 14-9. The angle formed by the ato in a water molecule.



balance in electrical charge makes the water molecule a *dipolar molecule*, which will orient itself in an electrical field like a magnet in a magnetic field. Water molecules may even align with each other because of the strong electrical attractions and repulsions between the molecules.

Water molecules, however, attract each other much more strongly than is indicated by their dipolar nature alone. If we examine a model of the water molecule, as in Fig. 14-10, we find that the hydrogen atoms produce two small protusions on the oxygen atom. The electrons shared between the hydrogen and oxygen atoms are held more closely to the oxygen atom because of the much greater positive charge on the nucleus of the oxygen atom (eight protons) than on the hydrogen atom (only one proton). The pair of electrons between hydrogen and oxygen are not shared equally, but are something like this, H :O, that is, the pair of electrons are actually closer to the oxygen atom than to the hydrogen atom. The hydrogen atom is left as an ion—as a bare nucleus of one proton—with the result that the two hydrogen atoms in the water molecule tend to become positively charged points and the oxygen atom tends to become negatively charged (Fig. 14-10b). Now we can begin to see how the positive hydrogen of one molecule is pulled toward the negative oxygen of another molecule (Fig. 14-10c) in the attraction called the *hydrogen bond*.

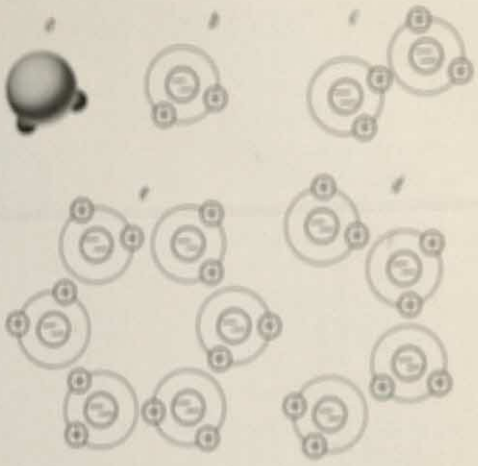


Fig. 14-10. The unusual properties of water are explained by the formation of water molecules associated with each other in small groups.

This hydrogen bond explains the unusual properties of water mentioned above. At ordinary room temperatures, three or four water molecules associate with one another through hydrogen bonds (Fig. 14-10d), so that when we drink a glass of water we are not drinking simple molecules of H_2O but agglomerates or clusters of water molecules, such as $(H_2O)_3$ and $(H_2O)_4$. The molecular weight of water, consequently, is in fact much greater than the 18 indicated by its formula. Water, as a result, acts abnormally when compared to the similar compounds in Fig. 14-11 that do not ordinarily form hydrogen bonds.

When water freezes, the molecules form even more hydrogen bonds than they do in liquid water. The angle of 105° between the hydrogen atoms in the water molecule requires a combination of six water molecules in order to produce a closed ring (Fig. 14-10e). Since there are two more points, one positive and one negative, on each water molecule in this ring that will become attached to still other water molecules, a complex network of six membered rings is formed in the ice crystal, which is indeed a beautiful thing (Fig. 14-12). The hexagonal arrangement of the water molecules in ice is responsible for the intricate hexagonal patterns that we often observe when we examine snowflakes.

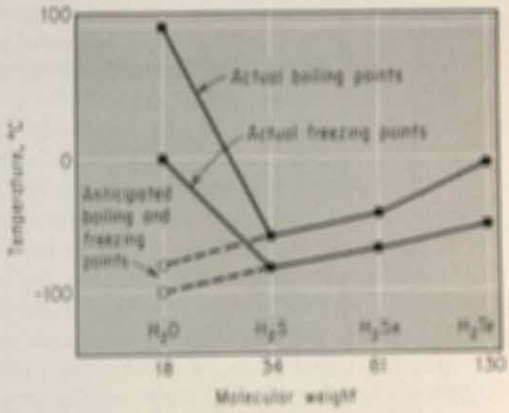


Fig. 14-11. A comparison of the boiling and melting points of compounds similar to water.

Since in Fig. 14-10e the hexagonal structure of the ice crystal contains an empty space in the center and since there is no corresponding space in liquid water, we may conclude that this difference explains why water molecules occupy 9 per cent more space as ice crystals than they do as liquid

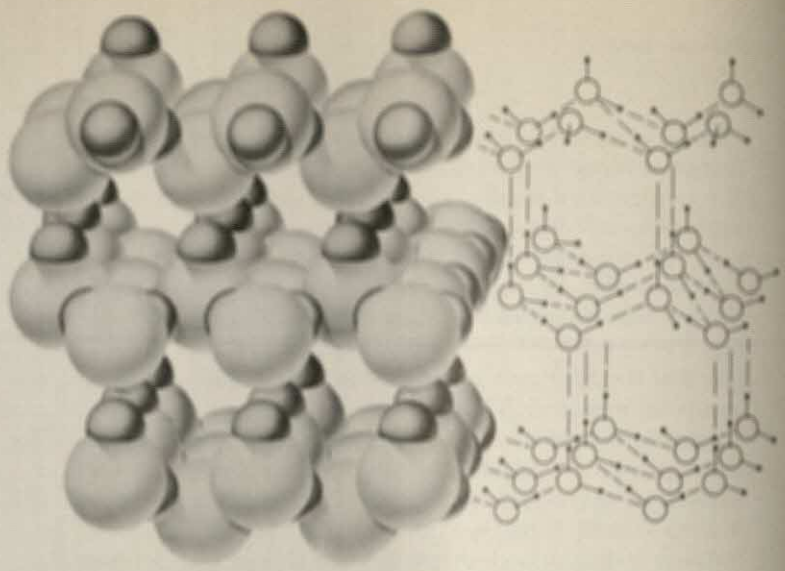


Fig. 14-12. The structure of the ice crystal.

molecules. When ice is melted, the contraction results from the fact that fragments of the ice crystal (as in Fig. 14-10d) fill in the space occupied by the original ice crystal. The maximum packing or density of water molecules occurs at a temperature of 3.98°C. At higher temperatures, the clusters of water molecules are increasingly broken up until just below the boiling point when they are mostly of the single H_2O variety. These unusual properties of water that result from the hydrogen bonds cause many natural processes that otherwise would not operate: the action of water upon freezing and thawing in degrading rocks and minerals, the evaporation and condensation of water into clouds, the behavior of water in lakes, streams and the ocean, and many other processes. The importance of water in these natural processes will be discussed in Chapters 23 and 24.

The hydrogen bond is also responsible for many other processes. The bond that forms between hydrogen and nitrogen, for instance, is thought to account for some of the important properties of proteins. It is likely that muscle protein is composed of long molecules that are coiled up into helixes which contract and extend during muscular activity. As these coils are compressed and stretched there must be certain forces that hold these coils in place, and since these bonds can be made and broken without disrupting the composition of the protein itself, the shape of the protein coil can be repeatedly changed without as a consequence destroying the protein itself.

There are other kinds of dipole bonds in addition to hydrogen bonds. An ion such as the iodide ion, I^-, may induce a dipole in what was otherwise a nonpolar molecule. This gives rise to the reaction:

$$I^- + I_2 \rightarrow I_3^-$$

and the I_3^- accounts for the fact that iodine is only slightly soluble in water but very soluble in a solution of potassium iodide. The iodide ions in the KI solution react with the I_2 molecules by means of ion-dipole bonds.

Another aspect of chemical bonds arises through a consideration of those chemical reactions that produce changes in the valence numbers of certain elements. This topic, called oxidation and reduction, will be examined in the next section.

14-8 Oxidation and Reduction

When we roll another log on the fire, we are providing more fuel for a reaction that goes by the name of oxidation-reduction. Whether we burn wood, coal, oil, or gas to heat our house, a chemical change takes place in the furnace that involves another important aspect of valence. Let us examine a few examples of this kind of process.

The reaction that occurs when charcoal or coke is burned is shown in the equation:

$$C + O_2 \rightarrow CO_2$$

When we burn natural gas, the reaction is somewhat more complex:

$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$$

And when gasoline, which is a mixture of compounds, burns in our automobile engines, a still more complex reaction occurs. One of these compounds, octane, burns according to the equation:

$$2C_8H_{18} + 25O_2 \rightarrow 16CO_2 + 18H_2O$$

In the burning of most common fuels, oxygen combines with the fuel to produce carbon dioxide and water. Rusting is another process in which oxygen takes an important part. When iron, for example, is exposed to the air, it turns into iron oxide:

$$4Fe + 3O_2 \rightarrow 2Fe_2O_3$$

Since all these reactions involve oxygen, they are called oxidation reactions.

Other reactions similar to those above, however, can be carried out without oxygen. Carbon, for instance, reacts with sulfur to form carbon disulfide, a substance widely used in industry:

$$C + 2S \rightarrow CS_2$$

Iron combines with chlorine, forming iron chloride:

$$2Fe + 3Cl_2 \rightarrow 2FeCl_3$$

The fact that carbon exhibits the same valence in carbon disulfide as it does in carbon dioxide and that iron has the same valence in iron oxide and iron chloride suggests that something other than the oxygen is the crucial factor here. Even though oxygen may not be involved in a chemical reaction of this kind, we have come to call all these reactions *oxidation*, and, accordingly, we say that any time a metal such as iron combines with another substance oxidation oc-

curs. On the other hand, when a metal compound such as an oxide or chloride is decomposed, we say that the metal has undergone *reduction*. Since oxidation and reduction always occur simultaneously, let us see if we can interpret these reactions in terms of our modern ideas of valence and electronic structure.

Oxidation is defined as the loss of electrons, *reduction* as the gain of electrons. When we burn sodium in chlorine, the reaction converts the sodium atom into a sodium ion with a valence of +1.

$$2Na + Cl_2 \rightarrow 2Na^+ + 2Cl^-$$

Suppose we write this equation in two parts, one part dealing with the change that occurs in the sodium and the other part dealing with the chlorine:

$$2Na \rightarrow 2Na^+ + 2 \text{ electrons} \qquad \text{(oxidation)}$$
$$Cl_2 + 2 \text{ electrons} \rightarrow 2Cl^- \qquad \text{(reduction)}$$

The sodium atom in this example of oxidation must lose one electron from its outer valence level to achieve electronic stability, and each chlorine atom gains an electron in a reduction process. Thus, when iron rusts—that is, when it reacts with oxygen—the iron becomes oxidized and the oxygen becomes reduced.

The loss-gain concept of oxidation-reduction explains why these two processes occur simultaneously. For a chemical reaction to occur, one substance must be ready to receive the electrons of another substance. In the electrolysis reactions of Faraday in Fig. 14-13, the oxidation-reduction process works like this: The passage of an electrical current through the molten sodium chloride breaks up the compound into its free elements by restoring the electrons in the chloride ions to the sodium ions. The net effect is oxidation of the chloride ion and reduction of the sodium ion; for each atom of chlorine formed there is one atom of sodium formed.

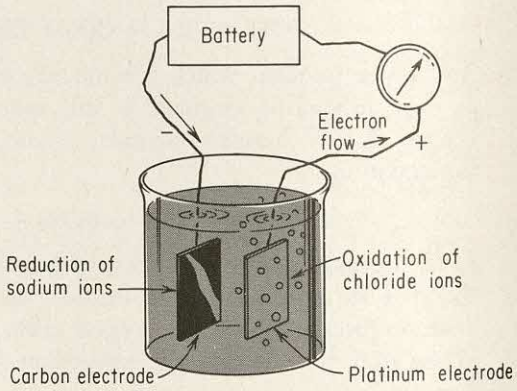


Fig. 14-13. Electrolysis of molten sodium chloride, showing the oxidation of the chloride ion and the reduction of the sodium ion.

The oxidation-reduction process is the key that unlocks metals from their impurities. Most metals, particularly the more reactive ones, occur naturally in the oxidized state: Iron, aluminum, and tin ores are essentially oxides of these metals. To obtain the free element from these ores, electrons must be restored to the metal ions to change them into metal atoms. Since we are primarily concerned with the metal, we usually call this process reduction, although there will necessarily be an oxidation reaction that accompanies it.

The two common methods of restoring electrons to metal ions are: electrolysis and chemical reduction. Aluminum is usually obtained from its oxide by electrolyzing a solution of aluminum oxide in molten cryolite, a sodium-aluminum fluoride salt. Electrons are restored to the aluminum ion at the cathode, and electrons are taken away from the oxygen at the anode. Iron, on the other hand, is usually obtained from its oxide by reducing it, not with an electrical current,

but with carbon, which at high temperatures combines with the oxygen in the iron oxide. The carbon undergoes oxidation and the iron undergoes reduction:

$$Fe_2O_3 + 3C \rightarrow 2Fe + 3CO$$

The reduction of iron oxide is carried out in huge blast furnaces that provide us with the iron and steel our modern civilization requires. In a similar way, tin is obtained from tin oxide by reducing it with carbon:

$$SnO_2 + C \rightarrow Sn + CO_2$$

Here, as the tin is reduced, the carbon is oxidized to its oxide, and, since the oxygen in this reaction keeps its same valence, it undergoes neither oxidation nor reduction.

Corrosion of metals is the reverse of the processes described above. Iron, when exposed to the atmosphere, returns to its more stable condition by oxidizing to rust. After aluminum, magnesium, zinc, and many other metals are freed from their ores, they do not long remain bright and untarnished but undergo oxidation into compounds similar to those that occur naturally. Even those metals, such as copper, that are well down in the reactivity series slowly corrode. A copper penny, heated in a bunsen flame, will soon form a red coating, cuprous oxide, Cu_2O, and then a black coating, cupric oxide, CuO, as more oxygen reacts with it. These two compounds represent two stages in the return of copper metal to its most stable valence in nature, +2.

Oxidation and reduction occur in many natural processes. Iron near the surface of the earth is oxidized to a valence of +3, but as we go below the earth's crust and into the warmer parts of the earth's interior, it is reduced to a valence of +2 or even zero (free iron). This change in valence is revealed by rocks, freshly exposed by erosion or volcanic activity, that are soon oxidized by the atmosphere into red, yellow, brown, green, or black colored rocks. Although water affects these colors to some extent, most of the color of rocks is caused by iron undergoing various degrees of oxidation. The presence of iron compounds in the rocky walls of the Grand Canyon of the Colorado River produces the brilliant hues that create that breathtaking spectacle, and the yellow iron compounds in the rock along the Yellowstone Canyon account for the name of this famous beauty spot in northwestern Wyoming. Hardly a mountainous state is without its favorite "Red Mountain," because iron, even in small quantities, is a very colorful element.

14-9 Summary

The atomic radii of the elements are periodic properties, and the largest atoms in a period lose their outer electrons, as measured by their ionization potential, more easily than do the smaller atoms. The smaller atoms tend to gain or

share electrons, and the number of electrons gained, lost, or contributed to a shared pair of electrons is called the valence of the element.

Faraday's laws of electrolysis show the relationship between the amount of electrical current and the number of chemical equivalents that undergo chemical change in electrolysis.

Electrovalence is the transfer of electrons from one element to another, producing ions in the process, and thus forming an electrovalent compound.

Covalence is the sharing of electrons between two atoms, with each atom furnishing an electron, and thus forming a chemical bond that holds these atoms together in a molecule.

Coordinate valence is the sharing of electrons between two atoms where one atom supplies both electrons for the bond. In both covalence and coordinate valence, the bond consists of the mutual sharing of one or more pairs of electrons by the two atoms involved in the bond.

The hydrogen bond is an electrostatic bond of a unique nature that arises from the imbalance of electrical charges in molecules that contain hydrogen bonded with oxygen, nitrogen, or fluorine. These special bonds account for many of the unusual properties of water as well as for the behavior of many complex substances, such as proteins.

Radicals are groups of atoms that act as a single atom or ion in many chemical reactions.

Oxidation is the loss of electrons, and reduction is the gain of electrons. These reactions occur simultaneously in all chemical reactions whenever there is a change in valence.

EXERCISES

A. KEY TERMS TO REMEMBER

atomic radii
coordinate valence
coordination number
covalence
dipolar molecule
electrolysis
electron affinity
electronegativity
electrovalence
equivalent weight
faraday
Faraday's laws
hydrogen bond
ion
ionization potential
oxidation
radical
reduction
valence

B. QUESTIONS ABOUT CHEMICAL BONDS

1. Draw diagrams of the following ions, showing the electrons that are lost or gained by the original atom in each case: Li^+, Na^+, Ca^{++}, Mg^{++}, Al^{+++}, Cl^-, F^-, Br^-, O^{--}, S^{--}, and P^{---}.

2. Arrange the following elements in decreasing order according to their ionization potentials: Na, Mg, Al, Cl, P, Si.

3. Which is the larger in each of the following pairs: Na or Na^+, Cl or Cl^-, S or S^{--}, Al or Al^{+++}?

4. On the basis of electronegativity, which of the following are bonded primarily by electrovalence or ionic valence, which by covalence: LiCl, LiH, CaO, SiO_2, AlP, CH_4, and H_2O?

5. Make electronic diagrams of the following electrovalent compounds, showing the electrons that are transferred in forming the molecule from the atoms: LiF, MgF_2, AlF_3, $CaCl_2$, CaO, and Ca_3P_2.

6. Show the electronic structure of the following covalent molecules: H_2, F_2, Cl_2, CH_4, CCl_4, CO_2, NH_3, and H_2S.

7. Write formulas based on the valence of the radicals in each of the following: sodium nitrate, potassium sulfate, calcium sulfate, magnesium nitrate, silver nitrate, aluminum phosphate, ammonium sulfide, ammonium carbonate, ammonium nitrate, and ammonium phosphate.

8. Explain why a radical such as the sulfate ion does not occur free in nature, but is associated with an ion of positive charge, such as calcium.

9. Explain why oxidation and reduction reactions always accompany each other.

10. In the following reactions, point out the elements that undergo oxidation and the ones that undergo reduction:

$2Zn + O_2 \rightarrow 2ZnO$
$2H_2 + O_2 \rightarrow 2H_2O$
$2Ca + O_2 \rightarrow 2CaO$
$2Na + Cl_2 \rightarrow 2NaCl$
$Fe_2O_3 + 3H_2 \rightarrow 2Fe + 3H_2O$
$Cu + Cl_2 \rightarrow CuCl_2$
$2Au_2O_3 \rightarrow 4Au + 3O_2$
$ZnO + C \rightarrow Zn + CO$

C. PROBLEMS ABOUT CHEMICAL BONDS

1. If 1000 coulombs are passed through an electrolysis cell where the reaction $Ag^+ + e^- \rightarrow Ag^\circ$ occurs, what weight of silver is formed at one of the electrodes?

2. Classify the following substances as polar, nonpolar, or ionic:
   a. CO_2, a linear molecule with the structure O=C=O.
   b. H_2O, an angular molecule.
   c. KCl
   d. CH_4, a tetrahedral molecule with carbon at the center.
   e. SO_2, an angular molecule.
   f. CH_3OH, a tetrahedral molecule with carbon at the center and an OH radical at one of the corners.

3. Show the compounds that can occur between Li, Na, and K and the elements F, Cl, and Br in the order of their stability as judged from Tables 14-4 and 14-5 and Fig. 14-2.

Fifteen

Ionic Substances

Atoms react with one another, as we have just seen, primarily by two processes—electrovalence and covalence. These two processes are fundamentally different, and we can therefore expect to find significant differences between the groups of substances formed by each of them.

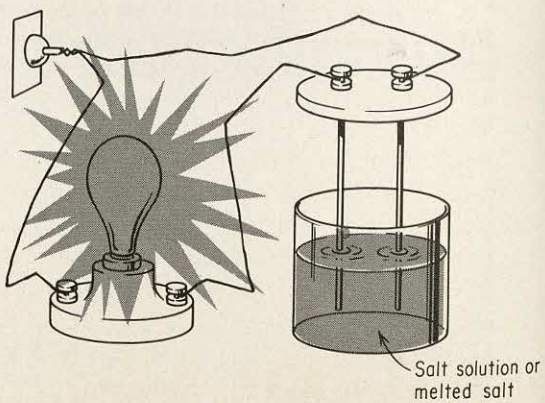
Since substances formed by electrovalence, or electron transfer, are composed of ions, they are sometimes called *ionic substances*. Although there are several types of ionic substances, depending upon their properties, it will simplify the study of them if we break them down into *oxides*, *acids*, *bases*, and *salts*. Before we examine these types of substances, however, let us acquaint ourselves with some of the general properties of ionic substances.

15-1 Electrolytes

All ionic substances are capable of conducting electricity, hence, we call these substances *electrolytes* (from the Greek, meaning to be decomposed by electricity). Ordinary salt (sodium chloride) does not conduct electricity because its ions are held rigidly in place in the salt crystal. However, if salt is melted we can pass an electrical current through it quite easily, as indicated by the glow of the lamp in Fig. 15-1. Sugar, on the other hand, will not conduct electricity even when it is melted, which is proof that sugar is not an ionic substance.

When we dissolve some salt in one container of water and some sugar in another, and again test their ability to conduct electricity, we will see that the salt solution con-

Fig. 15-1. An apparatus for testing the electrical conductivity of liquids and solutions.



ducts electricity readily but that the sugar solution fails to conduct any at all. Thus, we have an easy method for classifying a substance as either an electrolyte or a nonelectrolyte—we simply test its ability to carry electricity either when melted or in solution. From similar experiments we would find that baking soda (sodium bicarbonate), blue vitrol (copper sulfate) and alum (an aluminum salt) are electrolytes, but that gasoline, paraffin, and alcohol are nonelectrolytes.

In addition to the ability of ionic substances to carry electricity, they also exert an abnormal effect upon the freezing point of the liquids in which they are dissolved. For instance, if we dissolve one mole—6.02×10^{23} molecules—of sugar in one liter of water, the resulting sugar solution freezes at $-1.86°C$. That is, the presence of a mole of sugar molecules in a liter of water depresses the freezing point 1.86° below the freezing point of pure water. But, if a mole of salt is dissolved in a liter of water, the freezing point is abnormally depressed to $-3.48°C$, which is about twice as low as that caused by sugar. Apparently a mole of salt is about twice as effective as a mole of sugar in depressing the freezing point of water. The freezing point of other liquids in which these substances may be dissolved is similarly affected.

Another property of ionic substances, in contrast to covalent substances, is the high speed with which they react with each other. For example, solutions of sodium chloride and of silver nitrate, both electrolytes, react instantly when mixed together, producing a white, curdy precipitate of silver chloride. It is impossible to measure the time interval between the mixing of these two and the appearance of the white product. Sugar, on the other hand, takes part in reactions that are measurably slower; it can be burned in a few seconds, digested in a few minutes, or fermented in several days. Thus, in general, ionic substances react readily, but nonionic substances react slowly.

When these three properties of electrolytes were first recognized, scientists proposed various theories to explain them, but the first successful theory to account for the differences between electrolytes and nonelectrolytes was formulated by a Swedish chemist, Svante Arrhenius, in 1887. Arrhenius proposed that electrolytes in solution are dissociated into ions when an electrical current is passed through them. We know now, of course, that these ions are produced by a gain or loss of electrons from atoms or radicals, but the remarkable thing about Arrhenius' theory was that it was proposed many years before anything was known about atomic structure or electron gains and losses. When his theory was announced, it was eagerly tested and angrily discussed, and although its discoverer was ridiculed, even by his own professors, the ionization theory withstood the assault because it not only explained the facts known at that time but predicted many that were subsequently discovered.

As more refined experiments were performed, however, facts that did not exactly fit the theory were found, and, consequently, the original Arrhenius theory has had to be modified in order to account for what we know about electrolytes today. The three major properties of electrolytes that set them apart from nonelectrolytes—their ability to conduct electricity, to lower the freezing point of water, and to react quickly with each other—are explained by the modern theory in this way:

1. Electrolytes, either when melted or in solution, are composed of positive and negative ions. When an electrical current is passed through such a system, the positive ions move toward the cathode (negative pole) and the negative ions move toward the anode (positive pole) where they undergo oxidation or reduction by gaining or losing electrons. It is the movement of these charged particles, the ions, that is essentially an electrical current (see Fig. 14-13). Thus the Arrhenius theory explains how an electrolyte conducts electricity. The quantity of ions migrating and plating out at the electrodes has already been described by Faraday's laws. The degree of electrical conductivity of an electrolyte, dissolved or melted, is proportional to the freedom of motion of its ions—the freer they are the greater is their capacity to conduct electricity.

2. The abnormal effect of an electrolyte upon the freezing point of a liquid is also explained by the theory of electrolytes. The presence of dissolved material in a liquid interferes with the freezing process by making it necessary to cool the liquid below its ordinary freezing point before it begins to solidify. We saw above that a mole of salt depresses the freezing point almost twice as much as does a mole of sugar in the same quantity of water. The explanation for this difference is that the salt acts almost as if it were completely ionized. Thus, a mole of salt contains $2 \times 6.02 \times 10^{23}$ ions, while a mole of sugar contains only 6.02×10^{23} molecules. Apparently, ions and molecules are equally effective as depressants of the freezing point, and as a result a mole of salt interferes with the freezing process twice as much as a mole of sugar.

3. The instantaneous reaction rate of electrolytes is again explained by the presence of ions. Solutions of sodium chloride and silver nitrate, both electrolytes and both composed of ions, react rapidly with each other in a reaction that is usually written:

$$NaCl + AgNO_3 \rightarrow AgCl\downarrow + NaNO_3$$

All the substances in this reaction are ionic and, except for the silver chloride, very soluble. For the three substances that are both soluble and ionic it would be more nearly correct to show them in their ionic form:

$$Na^+ + Cl^- + Ag^+ + NO_3^- \rightarrow AgCl\downarrow + Na^+ + NO_3^-$$

Since nothing happens to the sodium ions and the nitrate

ions—they undergo no chemical change whatever—the equation can be simplified to show only the change that occurs between the silver ion and the chloride ion:

$$Ag^+ + Cl^- \rightarrow AgCl\downarrow$$

This equation relates all that happens chemically—the silver ions interact with the chloride ions to form insoluble silver chloride. Once we have arrived at the simple equation above for the reaction between silver nitrate and sodium chloride, we can readily see that this reaction is merely the interaction of positive and negative ions with each other. Once the two reactants are mixed, coulombic forces immediately draw ions of opposite charge together and the product is formed instantly. Hence, the theory of electrolytes, which holds that sodium chloride and silver nitrate are composed of ions, explains why ionic substances react instantly with each other.

4. Before concluding our consideration of the modern theory of electrolytes, we should examine one more point, namely, that there are noticeable exceptions that may complicate the picture unless the theory is more completely understood. For instance, we saw that one mole of sodium chloride does *not* lower the freezing point exactly twice as much as a mole of sugar. If the salt is completely dissociated into ions, they should be free to move independently of each other, and they should lower the freezing point exactly twice as much as the sugar does. Are we to assume, then, that the salt is not completely ionized? This would be the simplest assumption to make, and one that Arrhenius made, but modern study of this problem has given us a more adequate explanation, which is the most important change that has been made in the Arrhenius theory. The modern view is that the slight deviations in the behavior of electrolytes is due to the attractions that exist between ions of opposite charge. Let us look at the salt solution.

In sodium chloride solutions, all the sodium chloride molecules are ions, and each positive sodium ion, whether it is migrating to the negative pole in electrolysis or interfering with the freezing process, produces an electrical field about itself in accordance with Coulomb's law. Since there is an electrostatic attraction between each positive sodium ion and each negative chloride ion, each sodium ion attracts chloride ions and forms a sphere of chloride ions around it (Fig. 15-2). This whole cluster of ions tends to move along with the sodium ion wherever it goes and thus exerts a constant influence on it. Each chloride ion, in turn, builds up a sphere of sodium ions around it. The sphere of chloride ions around each sodium ion and the sphere of sodium ions around each chloride ion overlap and produce the interionic attractions that cause the sodium chloride to act as though it were not completely ionized. In very dilute solutions, where the distance between ions is great, these interionic attractions are relatively weak and produce only a slight drag on the ions, but in highly concentrated solutions, the ions are so closely

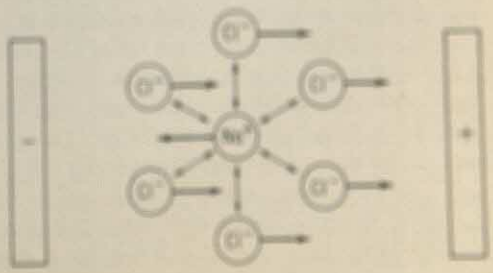


Fig. 15-2. The abnormal behavior of electrolytes is explained by the interionic attractions between ions of opposite charge.

spaced and exert such a large force of attraction on one another that the ions lose much of their ability to act independently.

In summary, the behavior of ordinary ionic substances is best explained by considering them as substances completely dissociated into ions. Coulombic forces between the ions do, however, cause some deviations from what we would expect of a substance that is entirely ionic, and these coulombic forces are most noticeable in solutions of high concentrations.

15-2 Solutions of Ionic Substances

Since many of the important properties of ionic substances become evident only when they are dissolved in water or other suitable liquids, let us examine briefly the characteristics of a solution. A *solution* is a mixture in which two or more substances are uniformly dispersed or distributed among one another on a molecular or ionic scale—that is, the molecules or ions of the substances involved are homogeneously distributed among each other in the solution. Because a solution is uniform and homogeneous it is the same throughout in color, density, taste, odor, and so forth. When salt is dissolved in water, we speak of the salt as the *solute* and the water as the *solvent*. Solutions may be dilute or concentrated, depending upon the relative amounts of solute and solvent. Molasses, a solution of sugars in water, is a concentrated solution, but ordinary drinking water, even though it may be very hard water, is only a dilute solution.

A solution is *unsaturated* when it is capable of dissolving still more of the solute; if a solution can dissolve no more solute when there is extra, undissolved solute present, it is said to be *saturated*. Most substances become more soluble when their temperatures are increased. If we make a saturated solution at a high temperature and then slowly cool it, for example, we may succeed in forming a solution, called a *supersaturated* solution, that may contain several times as much solute as a saturated solution of the same substance. A supersaturated solution is not a stable system; a slight jar or the addition of a speck of dust or a small crystal of the solute usually causes the excess solute to come out of solution. After the excess solute is removed from the solution, the resulting solution is merely a saturated solution.

The explanation of how an ionic substance dissolves is one of the more interesting achievements of scientific theory. Let us consider for a moment what happens when we add some ordinary salt to a container of water. You will recall that water has a dipole nature and that the sodium chloride crystal has an ionic structure. The water "wets" the surface of the salt crystal because the water molecules and salt ions attract each other. In other words, the water dipoles orient themselves about the edges and corners of the salt crystal (Fig. 15-3), and, since forces of attraction between

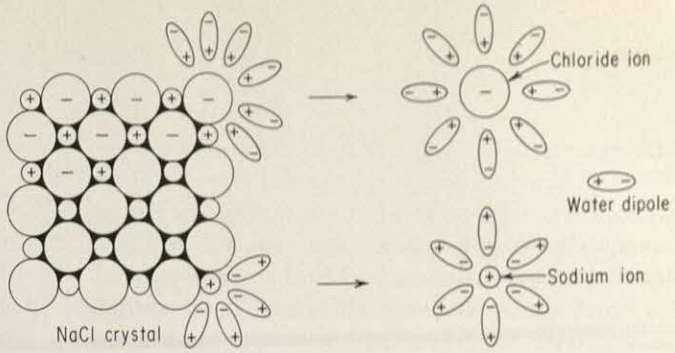


Fig. 15-3. The dissolving action of water upon sodium chloride.

the water dipoles and the ion in the crystal are sufficient to overcome the coulombic forces that hold the crystal together, the crystal of salt is gradually torn apart ion by ion. The ions that are dissolved are kept separate from one another by the numerous water molecules.

The extent to which a substance will dissolve in water is determined by the ability of the water dipoles to overcome the forces that hold the crystal together, and in this process, the valence of the ions, their size, and their ability to coordinate with the water molecule are the controlling factors. Almost all sodium salts are soluble in water to an appreciable extent because the sodium ion is readily surrounded by water dipoles and pulled away from the crystal. Many calcium salts, on the other hand, are insoluble or only slightly soluble because the valence of the calcium ion (+2) and its smaller size produce such strong coulombic forces in the crystals of calcium salts that the water dipoles are not capable of pulling the calcium ion out of the crystal.

Copper sulfide is another relatively insoluble substance. Only about fifty molecules of it will dissolve in a liter of water, because if a liter of water ever contains any more than about fifty copper ions and fifty sulfide ions at the same time, the excess ions will reunite to form a crystal of copper sulfide. Thus, a saturated solution of copper sulfide is indeed a dilute solution.

Nonionic substances, such as oil, wood, gasoline, cotton, and glass, are not soluble in water. A few nonionic substances such as alcohol and sugar, however, are soluble in water because their molecules are dipoles similar to water, and we often find a high degree of solubility among those substances that have similar dipolar properties. These dipolar properties arise in molecules that are unbalanced or unsymmetrical with respect to the atoms of which they are made. On the other hand, nonionic substances such as gasoline and paraffin, even though lacking dipolar properties, are soluble in each other because they are composed of molecules whose atomic arrangements are similar. In this case, the dissolving action is more or less a mechanical mixing process, with coulombic forces acting at a minimum. In short, we are safe in saying that like dissolves like—that ionic and dipolar substances are mutually soluble, and that nonionic, nondipolar substances are also mutually soluble. However, there is little tendency for substances in the first group to be soluble with substances in the second group.

15-3 Oxides

Oxygen combines with most of the elements—metals as well as nonmetals—to form oxides. Since oxygen is everywhere present in the atmosphere and in the upper parts of the earth's crust, oxides are very abundant. In its oxides, oxygen has a valence of −2 because it requires two electrons to reach electronic stability. The valence of oxygen is so dependable that it has long been used as a means of identifying the valence and equivalent weight of many other elements.

The stability of oxides varies depending upon the reactivity of the other element present in the oxide. Oxides of those metals at the top of the reactivity series are extremely stable; the oxides of magnesium and aluminum, for instance, are used as refractories or insulation lining in extremely high-temperature furnaces. Sand is silicon dioxide, another very stable oxide. Many oxides of the less reactive elements, on the other hand, are unstable and can be decomposed easily by heat. The oxide of mercury is of particular interest in this respect since it provided the first means of producing pure oxygen in the laboratory. Joseph Priestley, the English clergyman who pursued chemistry as a hobby, in 1774 produced oxygen by the simple process:

$$2HgO + \text{heat} \rightarrow 2Hg + O_2$$

He found that silver and gold oxides, upon heating, decompose into oxygen and the free metal.

Oxides are formed much more rapidly in pure oxygen than in atmospheric oxygen, which is diluted by nitrogen and other inactive gases. Priestley noted, for instance, that the pure oxygen he produced from mercury oxide caused a candle to burn with a remarkably brilliant and vigorous flame. This ability of pure oxygen to speed up the rate of burning is applied in several useful operations, one of which is cutting through iron and steel by means of an oxygen torch. The iron is first heated to redness with a regular flame, and then pure oxygen is forced onto the hot iron, causing the iron literally to burn up in the pure oxygen, forming iron oxide. Much of the pure oxygen produced commercially is used in oxygen cutting torches, which can easily burn through steel plates up to 12 inches thick.

15-4 Acids

For our present purposes, we can say that an *acid* is any substance that ionizes in solution to form hydrogen ions. It is also useful to classify those substances as acids that produce a distinctive effect upon many dyes. For example, litmus, a naturally occurring dye, is turned red by an acid. To acquaint you with this type of substance, several common acids and their formulas are listed:

Hydrochloric acid	HCl
Sulfuric acid	H_2SO_4
Nitric acid	HNO_3
Carbonic acid	H_2CO_3
Phosphoric acid	H_3PO_4

The sour taste of many foods is due to the presence of certain acids in them: vinegar is a dilute solution of acetic acid; rhubarb contains oxalic acid; and citrus fruits contain citric acid. Even a very dilute solution of an otherwise strong acid such as sulfuric acid tastes sour. A dilute solution of hydrochloric acid is an important factor in our digestive system; in fact, some people who cannot digest their food properly must supplement their natural supply of this acid by adding a small amount of hydrochloric acid to their diet.

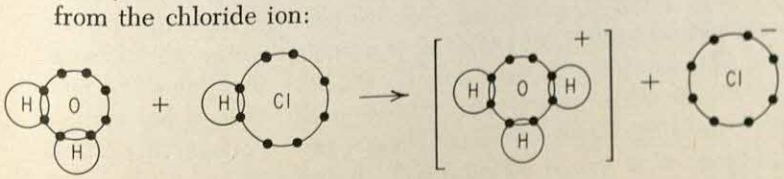
In the above list of acids you will note that there are two groups of acids—one containing only two elements, hydrogen and a nonmetal, and another containing three elements, hydrogen, oxygen, and a central nonmetallic element. There is an important relationship between acids and the position of these nonmetals in the periodic table. Indeed, to make a deliberate pun, the "acid test" of a nonmetal is the strength of the acid it forms when present in an acid—that is, as a general rule, the most active nonmetals form the strongest or most highly ionized acids. The elements in the upper right of the periodic table, consequently, form acids as binary compounds with hydrogen—HF, HCl, H_2S and so on—or as the central element in acids containing oxygen.

Now let us examine an acid in more detail in order to ascertain the factors that influence its acidic properties. Strong acids, such as hydrochloric acid, are highly ionized when in solution, as indicated by their ability to conduct electricity easily. Conventionally, the ionization of the hydrogen chloride molecule is shown in the equation:

$$HCl \rightarrow H^+ + Cl^-$$

Supposedly, it is the hydrogen ion and the chloride ion that account for the electrical conductivity of hydrochloric acid, and these ions are produced only when the acid is dissolved in water or other polar solvent. Pure hydrogen chloride alone is a nonelectrolyte, as are most other acids when not in solution.

Evidently, there is a reaction between the acid and the solvent, and, where the acid is present in an aqueous solution, the ionization process is actually brought about by the ability of the water molecule to pull the hydrogen ion away from the chloride ion:



which in formula form is:

$$H_2O + HCl \rightarrow H_3O^+ + Cl^-$$

It is probably the resulting hydronium ion, H_3O^+, that is responsible for the distinctive properties of an acid. An analysis of a solution of an acid once again reveals the unique behavior of water: water is an excellent solvent for acids because it is capable of causing an acid to undergo ionization.

Acids vary greatly in their ability to conduct electricity. Pure acids—that is, those that are not dissolved in water—are poor conductors of electricity because they are nonionic. All acids, however, form hydronium ions when they react with water and are thus able to conduct at least some electricity. Acids that ionize to a great extent in water are called *strong acids*, and they readily conduct electricity: some of these acids are hydrochloric, sulfuric, and nitric acids. Most acids do not ionize extensively in water and are therefore *weak acids;* their solutions are poor conductors of electricity because the water coordinates with only part of the hydrogen ions and the other molecules of the acid remain associated. Acetic acid is a weak acid: in vinegar less than 1 per cent of the molecules of acetic acid are ionized; the rest of the molecules remain as undissociated, nonionic molecules.

15-5 Bases

Bases are compounds that are the counterparts of acids. They are electrolytes when in solution, but they ionize to give a metallic ion and the hydroxyl ion (OH^-). They taste bitter or brackish and are capable of producing distinctive effects upon dyes, but opposite effects to those produced by acids. Litmus, for example, is turned blue by a base, and consequently this dye is commonly used to identify the presence of either an acid or a base. In the periodic table, the strongest bases are formed by the metals on the far left side of the table. In fact, the true measure of a metallic element is the strength of the base it forms in water—that is, the most metallic elements form the strongest bases.

Upon more careful examination of acids and bases, we find that the conventional idea of these substances may mislead us at times. For instance, a particular substance may act as an acid at one time and as a base at another. In order to arrive at a more dependable idea of these two kinds of chemicals many generalized concepts have been advanced in recent years that achieve a more useful definition of them. Let us examine briefly two of these modern concepts and apply them to common acids and bases.

According to Brønsted, a modern Danish chemist, all substances that can furnish a proton are acids, and all substances that can combine with a proton are bases. Thus, an acid is a *proton donor* and a base is a *proton acceptor*, and in equation form these definitions may be expressed:

$$\text{acid} \rightarrow H^+ + \text{base}$$

and:

$$\text{base} + H^+ \rightarrow \text{acid}$$

Note that the remaining part of an acid, after it gives up a proton, is a base, and that when a base reacts with a proton an acid is formed. The essential difference between an acid and a base is the proton—the acid has it, the base does not.

A second modern theory of acids, proposed by G. N. Lewis of the University of California, is that an acid is any substance that can form a covalent bond by *accepting* a pair of electrons. Conversely, a base is a substance that can *donate* a pair of electrons to form a covalent bond. This concept, although quite useful for nonaqueous as well as aqueous solutions, is somewhat beyond our interest in this book. Fortunately, we shall find the idea embodied in the hydrogen ion-hydroxyl ion view of acids and bases quite sufficient for an understanding of neutralization, which is the most common reaction of acids and bases in aqueous solutions. However, the two generalized theories of acids and bases—the Brønsted and the Lewis concepts—represent attempts to simplify the picture of acids and bases. Let us hold these ideas in mind as we turn to the interaction of acids and bases.

15-6 Neutralization of Acids and Bases

Since acids and bases have opposite natures—acids are proton donors and bases are proton acceptors—when we add an acid to a base a proton is transferred from the acid to the base. This transfer of the proton is called *neutralization* because the typical properties of both the acid and the base are mutually neutralized or changed. This process of proton transfer is similar to oxidation-reduction, in which electrons move from the oxidized to the reduced substance, and a new set of properties is produced.

Let us examine the neutralization reaction in more detail by considering a particular example. If we add an acid such as hydrochloric acid to a common base, such as sodium hydroxide, we observe an immediate reaction in which ordinary salt and water are formed:

$$\underset{\text{acid}}{HCl} + \underset{\text{base}}{NaOH} \rightarrow \underset{\text{salt}}{NaCl} + \underset{\text{water}}{H_2O}$$

Both in the laboratory and in nature, the neutralization of acids and bases commonly produces a salt and water. It is readily apparent that the water is formed by a proton donated by the acid combining with the hydroxyl in the base. The remaining ions—the chloride ion from the acid and the sodium ion from the base—are left over insofar as the neutralization reaction is concerned, and the combination of two such ions is called a salt. The Brønsted theory makes no provision for defining a salt, but in the conventional manner,

we usually call a *salt* a substance composed of a metal ion and a nonmetallic ion. The neutralization process is only one way in which a salt can be made.

That the neutralization of all bases by any acid is essentially the simple process above is shown when we measure the amount of energy involved. If equivalent amounts of any strong acid and any strong base are mixed together, the quantity of energy released by each of these neutralization reactions is, within experimental error, the same. The following list of reactions shows the amount of energy, in joules, accompanying the neutralization of various acids by various bases, all of which are strong electrolytes:

$$HCl + NaOH \rightarrow NaCl + H_2O + 57{,}500 \text{ joules}$$

$$HI + NaOH \rightarrow NaI + H_2O + 57{,}400 \text{ joules}$$

$$HCl + KOH \rightarrow KCl + H_2O + 57{,}600 \text{ joules}$$

$$HNO_3 + NaOH \rightarrow NaNO_3 + H_2O + 57{,}300 \text{ joules}$$

Since, within experimental error, the same amount of energy is released by each of these neutralization reactions, we may conclude that the same kind of reaction occurs in each case. In spite of the fact that different salts are formed in the four cases above, it is evident that the primary reaction in neutralization is the interaction of hydrogen and hydroxyl ions, and that when equivalent amounts of these ions react with each other we can expect the liberation, on the average, of 57,500 joules of energy.

For equivalent quantities of weak acids and weak bases, the amount of energy released is less than 57,500 joules. The reason for this becomes evident if we recall that weak acids, such as acetic acid, are only slightly ionized. Before these weak acids can be completely neutralized, they must become ionized, a process that uses up energy. If we add a small amount of base to acetic acid, the ionized acid is immediately neutralized (giving off energy), and some of the associated acids ionize (using up energy). As we continue to add base, the acid is both neutralized and further ionized, with the result that the total neutralization process involves a net amount of energy that is less than that produced by neutralization alone. For the neutralization of acetic acid by sodium hydroxide, for instance, the energy of neutralization is 56,000 joules.

15-7 Salts

As we have seen in the previous section, a *salt* is a compound composed of a positive metal ion and a negative nonmetallic ion. Since salts constitute both a large and useful group of chemical substances, many of which may be familiar to us in everyday experience, let us first consider how they can be made. There are primarily three different methods by which a salt may be produced, two of which you are already acquainted with:

1. It, along with water, is produced in the neutralization of an acid and a base.
2. It is the compound that is formed between a metal and a nonmetal by electron transfer.
3. It is often the result of a reaction between a metallic oxide and a nonmetallic oxide. Calcium carbonate, for example, may be produced by this method:

$$\underset{\text{calcium oxide}}{CaO} + \underset{\text{carbon dioxide}}{CO_2} \rightarrow \underset{\text{calcium carbonate}}{CaCO_3}$$

Salts occur both in solid form and in solution. Sometimes, when a solution of a salt is evaporated, the residue of salt still contains a certain amount of water, held to it by coordination bonds (p. 271), and that salt is called a *hydrate.* Since there is a definite relationship between the amount of water and the amount of salt in these hydrates, these systems should, and do, obey the law of definite proportions. For example, blue vitriol always contains 36 per cent water of hydration, and epsom salt contains 51 per cent water of hydration. In the true sense of a chemical compound, these hydrated salts exhibit a definite composition. In writing the formulas for these salts, it is customary to show the presence of the water in the hydrated salt:

$BaCl_2 \cdot 2H_2O$	Barium chloride
$CaSO_4 \cdot 2H_2O$	Calcium sulfate (gypsum)
$CuSO_4 \cdot 5H_2O$	Copper sulfate (blue vitriol)
$MgSO_4 \cdot 7H_2O$	Magnesium sulfate (epsom salt)
$KAl(SO_4)_2 \cdot 12H_2O$	Potassium aluminum sulfate (alum)
$Na_2CO_3 \cdot 10H_2O$	Sodium carbonate (washing soda)
$Na_2S_2O_3 \cdot H_2O$	Sodium thiosulfate (photographer's hypo)

Ordinary table salt, NaCl, however, along with many other salts, does not contain coordinated water, and even the hydrated salts may lose their coordinated water if sufficiently warmed. Such dehydrated salts are called *anhydrous salts.*

In addition to the salts listed above, some salts—called *acid salts*—are produced by the partial neutralization of an acid with a base. Every day we are likely to encounter some of these, one of which, baking soda (sodium acid carbonate, or sodium bicarbonate) results from the neutralization of carbonic acid with sodium hydroxide:

$$NaOH + H_2CO_3 \rightarrow NaHCO_3 + H_2O$$

Other common acid salts are:

$KHC_4H_4O_6$	Potassium acid tartrate (cream of tartar)
$NaHSO_4$	Sodium acid sulfate (sodium bisulfate)

When we reverse the process and partially neutralize a base by an acid, we produce a *basic salt:*

$Cu_2CO_3(OH)_2$	Basic copper carbonate (malachite, patina)
$Pb_3CO_3(OH)_4$	Basic lead carbonate (white lead)

Salts are a common type of compound. It has been estimated that over 30,000 different salts are known, and their composition and many of their properties as well. From A to Z, from aluminum to zirconium, the various metals supply the positive metal ion, and among the nonmetals and combinations of them there are at least 500 negative nonmetallic ions. It is evident that there are many possible combinations, using these positive and negative ions. Salts range from such simple and common ones as sodium chloride and calcium carbonate (the abrasive in your toothpaste) to such complex ones as antimonyl potassium tartrate hemihydrate (tartar emetic) and ferric ferrocyanide (Prussian blue, used as laundry blueing). Most of these many salts have little or no practical use, yet they constitute the achievement of many chemists who have pursued the strong human urge to create something new. And occasionally, surprisingly enough, one of these rare substances is pushed into prominence by a sudden discovery of its potential usefulness. Such has been the case with sodium fluoride (reduces dental decay) and zirconium oxide (protects against poison ivy and poison oak).

15-8 The pH of Water

The behavior of acids, bases, and salts is greatly influenced by the acid-base properties of water. To consider these types of ionic substances without considering the ionic properties of water would be to leave out some of the most crucial aspects of ionization. Up to now we have pictured water as primarily a covalent substance: The two hydrogen atoms are bonded by electron sharing to the central oxygen atom in the water molecule. But we cannot always draw a sharp line between covalent and electrovalent substances. Although water acts mostly as a covalent substance, it also has a certain amount of ionic character, and if we study the electrical conductivity of water with a sufficiently sensitive apparatus, we will find that water ionizes to a slight extent according to the equation:

$$H_2O \rightarrow H^+ + OH^-$$

A liter of pure water at room temperature contains 1×10^{-7} moles of hydrogen ions and an equal quantity of hydroxyl ions. This means that in a liter of pure water there are 1×10^{-7} times 6.02×10^{23} or approximately 6×10^{16} hydrogen ions and a similar quantity of hydroxyl ions, which represents less than 0.0000002 per cent ionization, but is enough to affect the properties of water.

The number of ions in pure water is insufficient for it to be an electrolyte, and since it contains an equal number of hydrogen ions and hydroxyl ions, it is neither an acid nor a base. When an acid is added to water the hydrogen ion concentration is increased, and at the same time the hydroxyl ion concentration is decreased, but never becomes

zero. An important relationship—called the *ion product of water*—exists between the relative amounts of these two ions that are contained in pure water or in solutions of either acids or bases in water. This relationship is based on the product of the molar concentrations of the hydrogen and the hyroxyl ions at room temperature, which is 10^{-14}.

In pure water the concentrations of hydrogen and hydroxyl ions are equal:

$$C_{H^+} \times C_{OH^-} = 1 \times 10^{-14}$$

Thus, since water is neutral:

$$C_{H^+} = 1 \times 10^{-7}$$

and:

$$C_{OH^-} = 1 \times 10^{-7}$$

Suppose we add enough acid to raise the hydrogen ion concentration to 1 mole per liter. Then:

$$C_{H^+} = 1$$

and:

$$C_{OH^-} = 1 \times 10^{-14}$$

Or, if we add a base to water to bring the hydroxyl ion concentration up to 0.1 mole per liter:

$$C_{OH^-} = 1 \times 10^{-1}$$

and:

$$C_{H^+} = 1 \times 10^{-13}$$

Since water is of strategic importance in many laboratory and natural processes and also in living organisms, scientists have devised a more convenient way to express its hydrogen ion concentration than that used above. By using a so-called *pH unit*, which is the negative exponent of the hydrogen

TABLE 15-1 THE pH OF ACIDS, BASES, AND NEUTRAL SYSTEMS

Hydrogen ion concentration (moles per liter)	*pH*		*Examples*
1 or 10^0	0	strongly acidic	
10^{-1}	1		
10^{-2}	2		
10^{-3}	3		gastric juice
10^{-4}	4		soda water (4.5)
10^{-5}	5	slightly acidic	
10^{-6}	6		
10^{-7}	7	neutral	water, blood (7.4)
10^{-8}	8		sea water (8.2)
10^{-9}	9	slightly basic	
10^{-10}	10		milk of magnesia (10.5)
10^{-11}	11		
10^{-12}	12		
10^{-13}	13		
10^{-14}	14	strongly basic	

ion concentration of a water solution, this concentration can be quickly designated. The relationship between pH units and acidic, basic, and neutral systems is shown in Table 15-1.

These pH units are widely used in medicine, agriculture, industry, scientific research—indeed wherever water solutions are studied. In medical examinations, for instance, the technician measures the pH of the blood, the urine, the gastric juices, and other body fluids in order to make a complete diagnosis of the patient. In agriculture, the pH of the soil is of primary importance in determining its fertility.

15-9 Hydrolysis

In addition to the unusual ability of water to dissolve many ionic substances, water has the unique property of being able to react chemically with many of these substances. To the small extent that it ionizes, water is capable of reacting quite vigorously with a substance, and when it does we say that the process called *hydrolysis* takes place.

Hydrolysis occurs most commonly between water and a salt that is made from either a weak acid or a weak base. For example, sodium carbonate hydrolyzes readily because this salt may be produced from sodium hydroxide (a strong base) and carbonic acid (a weak acid). When sodium carbonate is dissolved in water, a solution is obtained that has a very high pH, which means that it is a basic solution. This reaction may be explained by the equation for the hydrolysis reaction:

$$Na_2CO_3 + H_2O \rightarrow 2NaOH + H_2CO_3$$

Since the sodium hydroxide is much more ionized than the carbonic acid, the base is stronger than the acid. Thus, the sodium carbonate solution is basic.

If we view this reaction in the light of the hydrogen-hydroxyl relationship, we will find that, since sodium carbonate is a typical ionic substance, its ions react with water in the following manner (keep in mind that sodium carbonate exists as ions and that the ions themselves may or may not react with water, depending upon their acidic or basic properties):

$$Na_2CO_3 \rightarrow 2Na^+ + CO_3^{--}$$

$$2Na^+ + H_2O \rightarrow \text{(no chemical change)}$$

$$\frac{CO_3^{--} + H_2O \rightarrow H_2CO_3 + 2OH^-}{Na_2CO_3 + 2H_2O \rightarrow 2Na^+ + 2OH^- + H_2CO_3}$$

The sodium ion does not react appreciably with water, but the carbonate ion is a strong proton acceptor (a base) and therefore takes protons from water molecules, forming carbonic acid and converting the water molecules to hydroxyl ions. Thus, the net reaction that occurs when we add sodium

carbonate to water is the reaction of the carbonate ion with water:

$$CO_3^{--} + 2H_2O \rightarrow H_2CO_3 + 2OH^-$$

The solution of sodium carbonate is basic, then, because it contains considerable quantities of hydroxyl ions.

Solutions of salts that are made from a weak base and a strong acid are acidic, and solutions of salts made from an acid and a base of equal strength are neutral.

15-10 Water in Natural Processes

Water, because of its ability to dissolve various substances, invariably occurs in nature as the solvent in dilute solutions. Even rain water, the purest form of natural water, contains appreciable amounts of dissolved gases such as oxygen and carbon dioxide, and water pumped from the ground often contains dissolved salts that impart to it tastes, odors, and other characteristics that may or may not be desirable. The most concentrated solutions of natural waters are in inland seas such as Salt Lake in Utah and the Dead Sea in Jordan. All these natural waters are solutions, and their behavior is influenced by the fact that the dissolved solutes act as acids, bases, or salts. Let us examine the chemical action of natural waters upon their environments—a process called *chemical erosion.*

Rain water, as we noted above, is a dilute solution of dissolved gases, the most important one being carbon dioxide. Since carbon dioxide is a nonmetallic oxide, rain water is a slightly acidic solution, containing carbonic acid, which, in turn, ionizes to a small extent:

$$CO_2 + H_2O \rightarrow H_2CO_3 \rightarrow H^+ + HCO_3^-$$

Once the rain water falls to the earth it begins to attack the rock with which it comes in contact. One of the first rocks to be chemically altered is calcite or limestone ($CaCO_3$), large quantities of which are eroded away in this process. The presence of carbonic acid in the natural water causes it to react with the calcite crystals which consist of calcium ions and carbonate ions (Fig. 15-4). Carbonic acid is sufficiently ionized to furnish appreciable quantities of hydrogen ions which are attracted to the negative carbonate ions in the calcite crystal and which convert the carbonate ion (CO_3^{--}) to a bicarbonate ion (HCO_3^-). Since the salt, calcium bicarbonate, is quite soluble in water, the calcite crystal dissolves as it is converted to the bicarbonate salt.

The rain water, which has now seeped into the soil and dissolved quantities of calcium bicarbonate, is now called ground water, and it further dissolves water-soluble minerals, such as gypsum ($CaSO_4 \cdot 2H_2O$), various sodium salts, iron compounds, and so on. Such ground water, when pumped to the surface for domestic and industrial use, is given the de-

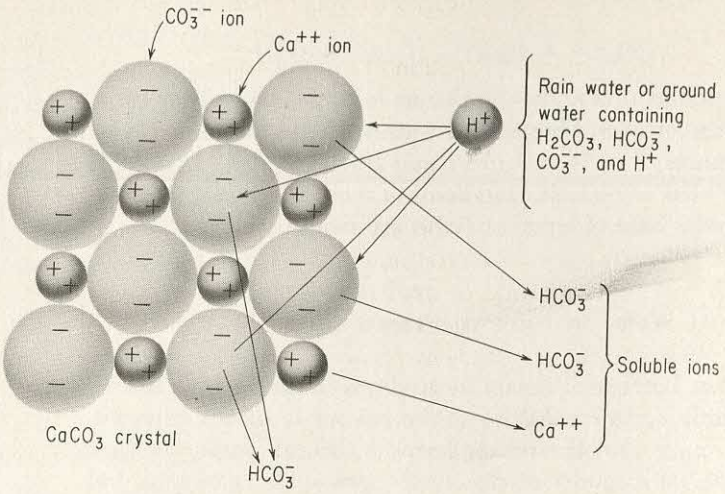


Fig. 15-4. A crystal of calcite being dissolved by water containing carbonic acid.

rogatory name, *hard water,* because the dissolved salts, even though present to a small extent, are undesirable chiefly for two reasons: One, the dissolved salts often form deposits of scale on the inside of water pipes and boilers and consequently cause serious damage; and, two, the presence of ions such as calcium and iron destroy the cleansing action of soap. The calcium ion converts soap (sodium oleate, for example) into an insoluble scum that accounts for the ring around the bathtub. This chemical reaction can be written:

$$\underset{\substack{\text{soluble}\\\text{(sodium oleate)}}}{2NaC_{18}H_{33}O_2} + \underset{\text{soluble}}{Ca^{++}} \rightarrow \underset{\text{soluble}}{2Na^+} + \underset{\substack{\text{insoluble scum}\\\text{(calcium oleate)}}}{Ca(C_{18}H_{33}O_2)_2}$$

It has been estimated that in a city of 40,000, a ton of soap would be destroyed daily if the typically hard water were not pretreated, or "softened," before use. Fortunately, it is relatively inexpensive to add water softeners that either remove the calcium and iron ions or render them harmless toward soap. In commercial operations, elaborate precautions are frequently taken to remove the solutes from natural waters in order to render the water safe for use.

15-11 Erosion of Limestone

In the last section we saw how calcium carbonate is dissolved by ground waters as these waters, acting as a dilute solution of carbonic acid, percolate down through the soil and rock layers at the earth's surface. In regions where limestone is common, the action of these carbonic acid solutions is so effective that enough rock may be dissolved to form tunnels and caverns. At a later stage, when these regions drain, these caverns become air-filled pockets or caves. These caves, found in many parts of the world, not only attest to the spectacular handiwork of the percolating solutions of carbonic acid, but provide a most challenging opportunity

for exploration. Only a relatively small portion of the caves in the world have been fully explored and developed for the sightseer.

At a still later stage in the action of ground waters in limestone regions, solutions of calcium bicarbonate seep through the roofs of these caves and form droplets from which a small amount of evaporation may occur. This evaporation both removes some of the dissolved carbon dioxide and increases the concentration of the calcium bicarbonate. Both of these effects lead to a supersaturated solution with respect to calcium carbonate, and a small amount of the calcium carbonate is precipitated. In time, a significant deposit of dripstone—called a *stalactite*—is produced. Water falling to the floor of the cave may form a similar deposit—called a *stalagmite*—and the combination of these two types of deposits often creates exquisite underground scenery (Fig. 15-5).

When large quantities of rock are dissolved by ground water, the overlying land may subside into the pockets

Fig. 15-5. The King's Palace in the Carlsbad Caverns of New Mexico. It has been estimated that these caverns were sixty million years in the making, but the stalactites and stalagmites were formed much more recently. (Courtesy National Park Service)

formed by the dissolving action. The resulting depressions are called *sinks*, and if they become filled with surface water, a *karst* lake appears. This type of landscape is common in the Cumberland plateau of Kentucky and in parts of France, Italy, and Yugoslavia.

Another outcome of the erosion of limestone and the accompanying solutions of calcium bicarbonate that permeate the upper layers of the earth's crust is the cementation of sand into *sandstone*. Sand grains are compacted and cemented together to form sandstone in regions near limestone, and it is probably the pressure and the temperature changes on the solutions of calcium bicarbonate that cause these solutions to become supersaturated with respect to calcium carbonate. The calcium carbonate that precipitates among the grains of sand serves to bind these grains together, thus forming sandstone.

15-12 Erosion of Granite and Other Rocks

The rate of water's attack upon granite and other rocks is much slower than its rate of attack on limestone. Nevertheless, over extended periods of time even the most resistant parts of the earth's crust are affected by water, and there are several ionic substances formed in these processes that are of interest at this point. Granite, for example, is a typical igneous rock that is slowly decomposed by the action of ground water. (It is called an *intrusive rock* because it is formed within the earth's crust as molten rock which then cools; in contrast, lava rocks are *extrusive rocks* because they are formed by the solidification of molten rock at the earth's surface.) Let us examine the chemical process that takes place.

Granite is essentially a mixture of three substances that can be visually examined in any ordinary piece of this kind of rock. Approximately half of it is feldspar, a quarter is silica or quartz (SiO_2), and the remainder, ferromagnesium silicates such as mica. *Feldspars* vary in composition: they are either sodium, potassium, or calcium aluminosilicates, the formulas of which may be shown as combinations of metallic and nonmetallic oxides:

Feldspars	*Chemical formula*
Orthoclase	$K_2O \cdot Al_2O_3 \cdot 6SiO_2$
Albite	$Na_2O \cdot Al_2O_3 \cdot 6SiO_2$
Anorthite	$CaO \cdot Al_2O_3 \cdot 2SiO_2$

Ground water attacks feldspars by acting both as an acid and as a hydrolyzing agent. Orthoclase, for instance, may be decomposed by water alone, or by the weakly acidic solution of CO_2 in water (carbonic acid). It is difficult to write adequate chemical equations describing the weathering proc-

ess, but the final products include silicic acid and clay, along with soluble ionic substances. With the complexity of these reactions in mind, the hydrolysis of orthoclase may be approximately represented by the equation:

$$K_2O{\cdot}Al_2O_3{\cdot}6SiO_2 + 11H_2O \rightarrow \text{(Intermediate products such as hydrous oxides and cations)} \downarrow$$
$$Al_2O_3{\cdot}2SiO_2{\cdot}2H_2O + 4H_4SiO_4 + 2K^+ + 2OH^-$$
kaolinite clay; silicic acid

Part of the soluble ionic products are carried away in solution by streams and rivers to the ocean, but a great part of it combines with clays and is used up by plants as soil nutrients before it gets to the sea. The silicic acid may remain in solution, but it frequently polymerizes (see p. 314) to large sized aggregates of chalcedony or opal. The silicic acid does not form sand, but the sand from the weathering of granite comes principally from the quartz in granite. Other types of feldspars are similarly attacked by ground waters to produce sodium ions and calcium ions, and, as we might expect, these products take part in natural processes similar to those involved in the decomposition of orthoclase.

The second component of granite—*quartz*—is only slightly attacked by water. Some of it is dissolved or carried off as fine, suspended particles, but larger particles are left behind as grains of sand. As we have already noted, a large number of these sand grains have been compacted and cemented together in the geological past, forming sandstone.

The third component of granite is the *ferromagnesium silicates,* and they undergo hydrolysis as ground water attacks them. These rocks contain iron, magnesium, and calcium as aluminosilicates, and they are decomposed into soluble salts, silica, clay, and occasionally into iron and magnesium compounds of industrial significance. The iron may eventually be found as red hematite, Fe_2O_3, an important iron ore, or as limonite, $2Fe_2O_3{\cdot}3H_2O$, a yellow, or sometimes brown, substance. All these reactions, though sometimes complex, are essentially ionic reactions.

Because of the wide variations in the composition of other kinds of rocks that are attacked by ground water, we can anticipate that a wide variety of chemical reactions occurs in this realm. However, not only does ample geological evidence exist for many of these reactions, but laboratory conditions can be set up to simulate the natural processes, and many of these complex reactions can thereby be studied more intimately. For instance, in the decomposition of volcanic rocks and in particular the ash that is erupted from volcanoes—called *volcanic ash*—the rocks and ash are believed to react almost instantly with water as they fall to earth. The product of this rather rapid reaction is a special kind of clay called *montmorillonite,* which is found in deposits near old volcanic regions.

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road

15-13 Summary

Electrolytes are characterized by their ability to conduct electricity, either when melted or in solution. They affect the freezing point of solutions to an abnormal extent, and they react instantaneously. These three fundamental properties of electrolytes are attributed to the ionic character of these substances and were first explained in the Arrhenius theory of ionization.

An important extension of the Arrhenius theory concerns the effect ions exert on one another—the so-called interionic forces that make them appear to be slightly associated.

A solution is a molecular or ionic dispersion of two or more components, which may be dilute or concentrated (relative terms), or saturated, unsaturated, and supersaturated.

Solubility of one substance in another follows the general rule of like dissolves like. Ionic and polar substances are soluble in polar solvents, and nonpolar substances are soluble in other nonpolar substances if they are similar in constitution and make-up.

Metal oxides dissolve in water to form bases, and nonmetallic oxides react with water to form acids. Bases contain the hydroxyl ion and acids contain a hydrogen ion. Modern concepts of acids and bases view an acid as a proton donor and a base as a proton acceptor.

An acid and a base neutralize each other and form a salt and water. The neutralization reaction is essentially an interaction between a hydrogen ion and a hydroxyl ion, forming water. The salt is composed of the metal ion of the base and the negative ion of the acid.

The pH of an aqueous solution is the negative exponent of the hydrogen ion concentration.

Hydrolysis is the decomposing action of water upon a substance. Salts are quite commonly hydrolyzed to corresponding acids and bases.

Water in nature is a dilute solution of carbonic acid and acts as an acid and hydrolyzing agent upon limestone, granite, and other rocks. The formation of caves, karst lakes, and hard water are directly attributed to the action of ground waters in dissolving limestone. Granite is converted by ground waters to clays, silicic acid, and various soluble salts.

EXERCISES

A. KEY WORDS TO REMEMBER

acid	hydrolysis	sink
base	ionization	solute
cavern	karst lake	solution
chemical erosion	neutralization	solvent
electrolyte	oxide	stalactite

feldspar	pH	stalagmite
granite	salt	supersaturated solution
ground water	saturated solution	unsaturated solution
hard water	silica	

B. QUESTIONS ABOUT IONIC STRUCTURE

1. List three major differences between electrolytes and nonelectrolytes.

2. Show how the theory of ionization explains the differences between electrolytes and nonelectrolytes.

3. If you were given a solution of a salt and a small crystal of the same salt, tell how you could test the solution to learn if it were saturated, unsaturated, or supersaturated.

4. When solutions of copper sulfate and sodium sulfide react, a precipitate of copper sulfide is formed. The other product, sodium sulfate, is still soluble. Rewrite the equation to show only those ions that are involved in the chemical reaction: $CuSO_4 + Na_2S \rightarrow CuS\downarrow + Na_2SO_4$.

5. Write chemical equations to show how oxygen is liberated by heating these compounds: HgO, $KClO_3$, Au_2O_3, Ag_2O.

6. Which of the following substances are acids, which are bases, and which are salts?

HNO_3	$CuSO_4$	H_2CO_3	$Al(OH)_3$	$CuCl_2$	$HBrO_3$
$Mg(OH)_2$	H_2SO_4	$CaCO_3$	Na_2CO_3	H_3PO_4	NH_4OH
$AlCl_3$	H_3BO_3	NaOH	KCl	$HClO_3$	NH_4Cl

7. Name the substances in the previous question.

8. Explain why the same amount of energy is released by the neutralization of any strong acid by any strong base.

9. Tell what acid and what base may be used to make each of these salts:

sodium chloride	potassium nitrate
sodium acetate	aluminum chloride
calcium sulfate	sodium borate
magnesium carbonate	sodium carbonate

10. Classify each of the following systems as either acidic, basic, or neutral:

(a) a solution of sodium chloride
(b) pure water
(c) a solution with pH 7.5
(d) a solution with pH 14
(e) a solution with pH 2
(f) a solution of sodium carbonate
(g) vinegar
(h) rain water

11. Explain how calcium carbonate is dissolved by ground water. Is it dangerous to use limestone as a building stone where rain water comes in contact with it?

12. If you visited a strange country and found many small lakes, how could you decide if these were karst lakes? What information would you need in addition to the evidence available from surface observations?

Sixteen

Covalent Substances

Now we turn to substances that are held together primarily by covalent bonds. In contrast to electrolytes, which are composed of ions, the properties of covalent compounds are determined by the behavior of discrete molecules that act much more independently of each other than do the ions in an ionic substance. The molecules in water, for instance, determine the properties of water, while, with ordinary salt it is the sodium ions and chloride ions—not molecules of sodium chloride—that determine the properties of salt. For this reason, we will find a whole set of properties of covalent substances—melting point, boiling point, and electrical conductivity—that are the result of weak forces between neighboring molecules, while similar properties for ionic substances will be radically different because there are strong electrical forces that act between the ions composing the substance.

In our study of covalent substances, we will first consider a few elements that form covalent bonds; then we will examine some of the common carbon compounds that are representative of covalency.

In contrast to the metallic bond (see p. 257) that accounts for the strength, ductility, and conductivity of metals, nonmetallic elements are usually held together by covalent bonds, and the difference in these bonds is largely responsible for the difference between metals and nonmetals. The covalent nonmetals are usually brittle, colored, and low in mechanical strength. Their melting points vary widely. At room temperature, some nonmetals are gaseous, some are liquid, and some are solid. As we examine a few of the more common elements that exist as covalent molecules, we shall learn more about these contrasting properties of the nonmetals.

16-1 Elementary Covalent Substances

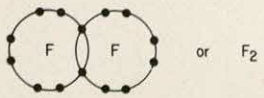
The hydrogen molecule (see p. 228) is an excellent example of a purely covalent bond. In it, the two hydrogen atoms share their single electrons, which form a bond that holds the two atoms together, and are fixed in space with respect to each other. Once the two hydrogen atoms are joined, they remain united until a vigorous reaction separates them. They may vibrate, and they may spin or rotate around a point midway between them, but they remain firmly attached to each other, and, until the molecule is broken apart, the average distance between their centers is a constant 0.74 A. The boiling point of hydrogen is extremely low—253° below zero Celsius—where the kinetic energy of the hydrogen molecules becomes small enough so that the extremely weak forces between the hydrogen molecules may draw these molecules into a liquid form.

Hydrogen has captivated the interest of scientists for many years. It was one of the first elementary gases to be discovered and made in the laboratory. Since hydrogen gas is easy to prepare—the hydrogen is displaced from an acid by a metal that is above hydrogen in the activity series—and since the gas is highly inflammable, its combustion properties were intensively studied by many early scientists. When hydrogen burns in air, or with oxygen, it is converted to water:

$$\underset{\text{hydrogen}}{2H_2} + \underset{\text{oxygen}}{O_2} \rightarrow \underset{\text{water}}{2H_2O}$$

Because of the strength of the hydrogen-oxygen bonds in water, hydrogen frequently reacts with oxygen even when it is already bonded to other elements. For example, at elevated temperatures, hydrogen removes the oxygen from iron oxide, forming water vapor and free iron. Many other metal oxides can be similarly converted to the free metal, and since this reduces the valence number of the metal it is called a *reduction reaction* (p. 279).

Several other elements are held together by covalent bonds in diatomic molecules. In Group VII of the periodic table, there are four covalent elements that we call the *halogens*. They are fluorine, F_2, chlorine, Cl_2, bromine, Br_2, and iodine, I_2. The covalent bonds in them are similar to one another. In fluorine, each atom originally has seven electrons in its outer energy level, and it readily shares one of its electrons with another fluorine atom. By completing the outer rings of both atoms, the two electrons form a covalent chemical bond between the atoms.



or F_2

Since all the other halogens have seven electrons in their outer electron levels, the molecules of these elements are bonded by covalence in the same way as the fluorine molecule.

Sulfur is an element that has some unusual characteristics in its covalent bonding. With atomic number 16, it has six electrons in its outer orbit and finds two more electrons by sharing two electrons in covalent bonds with other sulfur atoms—with more than one, because a single molecule of ordinary sulfur (a pale yellow, brittle substance) contains eight sulfur atoms (S_8). In place of a simple covalent bond, each sulfur atom is tied by a shared pair of electrons to two other sulfur atoms, for the sulfur molecule is a ring of eight sulfur atoms (Fig. 16-1a).

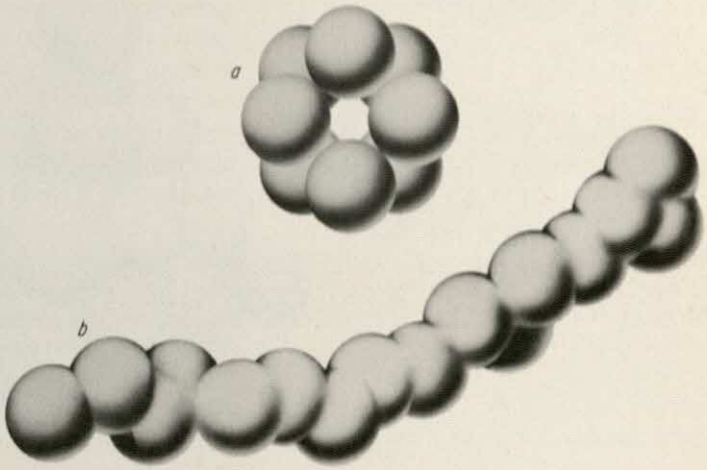


Fig. 16-1. The sulfur molecule, S_8, and the linear sulfur molecule.

Sulfur melts at approximately 120°C, and as it is further heated, the pale yellow color changes to a deep red color and the liquid becomes very viscous, so that it no longer flows as an ordinary liquid. This great change in the properties of sulfur is explained by the fact that the S_8 rings open up and form long chains of sulfur atoms (Fig. 16-1b). Hundreds of sulfur atoms are now attached to one another through covalent bonds, and as these long molecules become entwined with each other they account for the viscous nature of sulfur. When molten sulfur is quenched in cold water it forms a plastic, rubbery mass that is called *amorphous sulfur* which, according to recent X-ray studies, is still composed of long chain molecules but they are coiled up into helices with eight sulfur atoms in each turn of the helix. Upon standing, the helical structure gives way to individual S_8 rings and ordinary crystalline sulfur is re-formed.

No doubt the most interesting element that is held together with covalent bonds is carbon, which occurs in two different forms: diamond, the hardest substance known, and graphite, one of the softest substances known. Diamonds are often clear, brilliant, and highly refractive gems; graphite is black, slippery, and crumbly. The vast differences between diamond and graphite are caused by the differences in their crystal structure.

In the diamond structure, each carbon atom is tied by a

regular covalent bond to four neighboring carbon atoms (Fig. 16-2). Since each carbon atom is only 1.54 A from its neighbor, the bond in diamond is one of the shortest and strongest known. Although you might think that the unusual hardness of diamond is explained by its close-packed structure, the density of diamond is actually only half that of the tightest close-packed system; it is a relatively open structure, as is shown in Fig. 16-2. The hardness of diamond is

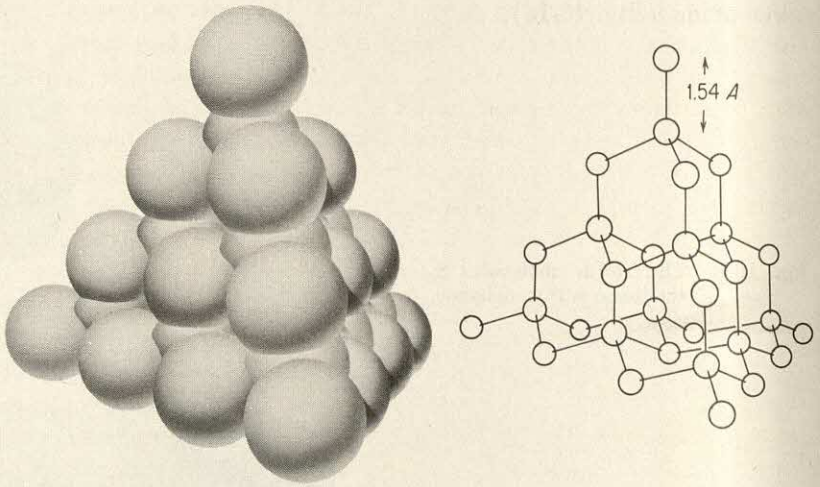


Fig. 16-2. The crystal structure of diamond.

caused by two factors: the strong carbon-carbon bonds in the crystal and their unique arrangement, which provides a very rigid network of atoms within the crystal. A diamond, in effect, is one huge molecule, since all the atoms in a single gem are held together in an unbroken covalent bond. The size of the molecule depends upon the size of the gem, and for this reason we cannot write a precise formula for the whole molecule. When all the atoms in a crystal are connected in this fashion, we call it a *giant molecule.*

The conditions under which diamonds are now made industrially provides a clue to the conditions that must have prevailed when natural diamonds were made. In the manufacture of diamonds a pressure of one and a half million pounds per square inch and a temperature above 2800°C—conditions that probably prevail at a depth of 240 miles below the earth's surface—are required in order to orient the carbon atoms into the diamond crystal. Synthetic diamonds are identical to natural diamonds, except that they are small and contain impurities that rule them out as gem quality diamonds.

Graphite consists of layers or sheets of carbon atoms. Although the atoms in these layers are only 1.42 A apart—closer to each other than are the carbon atoms in the diamond—they are exactly in a plane, which produces the smoothest surface known. The sheets are 3.41 A from each other and have little attraction or bonding between them (Fig. 16-3). Since the layers may slip over each other very

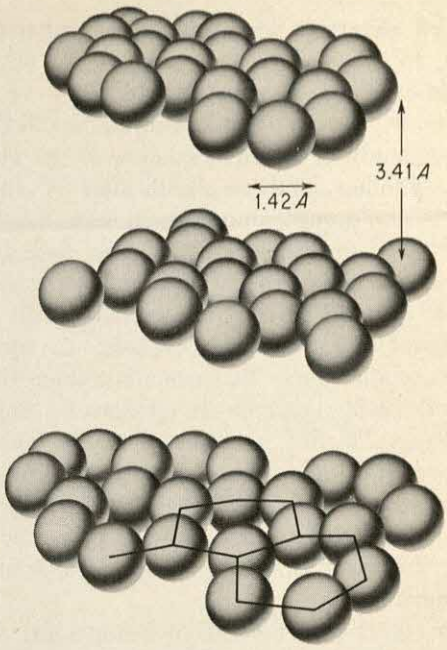


Fig. 16-3. The crystal structure of graphite.

easily, graphite is an extremely uncohesive, or soft, substance. Another quality of graphite is its ability to conduct electricity in a direction parallel to the layers but not at right angles to them. This indicates that graphite must have electrons that are free to migrate through the crystal. The diagram of the structure of graphite shows that each carbon atom has only three regular covalent bonds. The fourth electron forms another kind of bond in which the electrons may shift easily from one pair of atoms to another. The unstable nature of this fourth valence bond for each carbon atom accounts for the electrical properties of graphite. Since carbon is near the middle of the periodic table, it is not surprising to find that it has certain metallic properties, but nevertheless it is basically a nonmetal.

Silicon, an element directly below carbon in the periodic table, has many properties similar to carbon. It has a covalence of 4 and the structure of its crystal is identical to that of the diamond. It is considerably softer than diamond because its atoms are much larger than carbon atoms, and thus much further apart. The most important use of silicon is as a deoxidizer in the manufacture of high-grade steel: Silicon reacts vigorously with the dissolved oxygen in melted steel to form a very stable compound, silicon dioxide (which is insoluble in the steel), thus removing dissolved and combined oxygen from the steel.

As a rule, the larger the atoms in a crystal and the farther they are from each other, the weaker the covalent bond that holds them together. This is especially noticeable among the elements in Group IVA in the periodic table. Crystals of the elements below carbon and silicon become progressively softer. These elements commonly show a covalence of 4, and most of them form the diamond-type crystal, but the covalent bonds become so weak that in the case of tin the crystal hardly holds itself together. Tin in the diamond structure is actually in the form of a dust. Ordinarily, however, tin is more a metal than a nonmetal; we usually see it as a shiny, soft, malleable metal, frequently plated onto steel to make the common "tin" can. But at temperatures below 18°C, tin slowly changes into the nonmetallic structure of the diamond, and this transformation can be quite dramatic. Once, during a long, cold winter in northern Europe, a cathedral's organ pipes, afflicted with what is known as "tin disease," crumpled to dust. The tin, in short, changed from a metallic crystal to a nonmetallic, diamond-type crystal. When tin is heated to temperatures above 18°C, the transition is reversed. When tin is mixed with other metals or plated in a thin layer on other metals, such as in the ordinary tin can, tin shows very little tendency to change to the diamond form at low temperatures.

16-2 Organic and Inorganic Substances

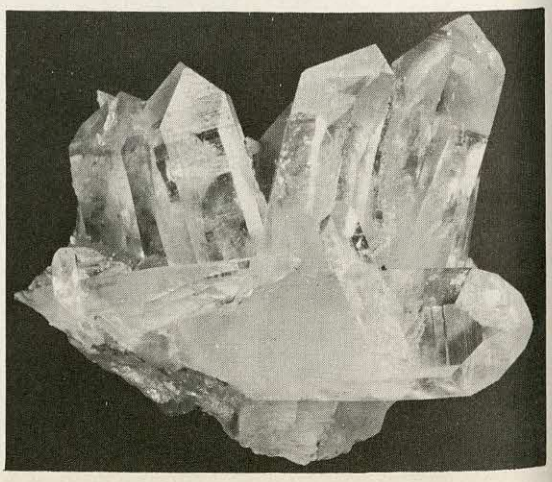
We have seen how substances may have primarily electrovalent or primarily covalent properties. Before going further in our study of covalent substances, we should briefly consider another kind of classification of substances. Carbon and silicon occupy unique positions in this classification: carbon is the central element in one class, silicon the central element in the other. At one time it was believed that compounds of carbon could be prepared only by a plant or an animal. Hence, the name "organic" was applied to the study of those substances that contain carbon. *Organic substances* include such common compounds as sugar, alcohol, and benzene. Possibly a million organic compounds are now known, and most of them have been synthesized in the laboratory. Although the idea that carbon componds must be derived through organic or living processes is now long obsolete, the study of carbon compounds is still called *organic chemistry.*

In contrast to the carbon compounds are all those substances that do not contain carbon—we call them *inorganic compounds.* The simpler inorganic compounds include many salts, acids, bases, and oxides, which were discussed in the last chapter. Among the more complex inorganic compounds are those that contain silicon, an element that most of us have never seen although, since most of the rocks of the earth's crust are complex silicates, we encounter many of its compounds every day. Let us now examine some of the important properties of this element as it exists in covalent, inorganic substances.

16-3 Silicon Compounds

Silicon commonly occurs in nature as the oxide, SiO_2. The most familiar form of silicon dioxide, or silica, is the quartz crystal (Fig. 16-4), which shows us that silica is not composed of simple SiO_2 molecules, but is rather a giant molecule, with each quartz crystal being a huge, single molecule.

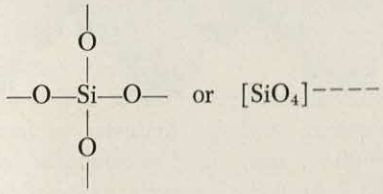
Fig. 16-4. Several giant molecules of SiO_2 in the form of quartz crystals. Courtesy Ward's Natural Science Establishment.



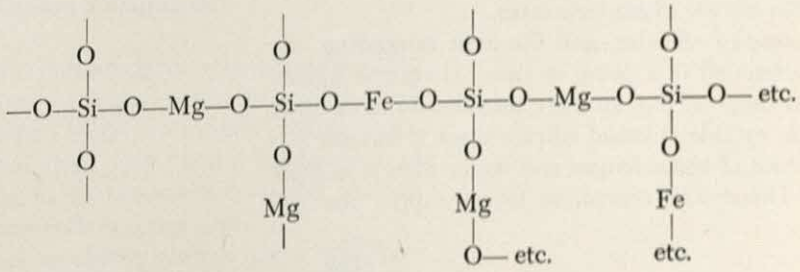
Silica possesses several unique properties. When melted it can be shaped into various pieces of laboratory ware much like glass. However, the high melting point of silica, about 1600°C, makes it especially useful for high-temperature experiments where ordinary glass, which melts or softens at 600–800°C, would be useless. Fused silica glass also has the ability to withstand sudden temperature changes. The shock of plunging a red-hot silica vessel into icewater, for instance, fails to crack the glass. Another useful property of silica glass is its transparency to ultraviolet light; most common glasses are almost opaque to ultraviolet light. Silica's extreme stability makes it chemically inert, and laboratory apparatus made of silica glass is attacked only by the most vigorous reagents.

The valence bonds in silica are all covalent bonds, and no more than one pair of electrons is ever shared between each pair of silicon and oxygen atoms. The arrangement of four oxygen atoms around each silicon atom reminds us of the arrangement of the carbon atoms in the diamond crystal. This so-called *tetrahedral arrangement* is an important aspect of the structure of all silicas, simple as well as complex. Silica contains uninterrupted chains of silicon-oxygen-silicon-oxygen bonds, and by introducing other constituents such as metal oxides into the silica, we can obtain complex silicates such as feldspar, mica, and clay. Let us examine some simple and complex silicates to see how they are related to the original silica structure.

Here is a single silicon atom surrounded by four oxygen atoms:

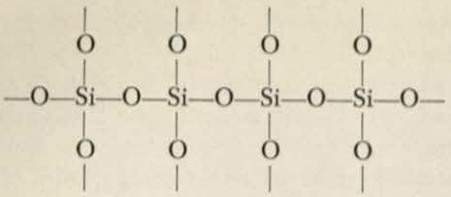


If four sodium ions are attached to this simple silicate radical, we have sodium silicate, Na_4SiO_4. A solution of sodium silicate, sometimes called water glass, serves as a cheap adhesive and as a fireproofing agent. Among the naturally occurring minerals that contain the SiO_4 radical are the olivines, a type of mineral of which garnet is an example. In these minerals, a divalent metal, magnesium or iron, is attached to the silicate group, and the Mg^{++} or the Fe^{++}, with its capacity to form two bonds, holds two silicate groups together. The result is a continuing network of silicate groups:



Since the similarity of the ionic radii of Mg^{++} and Fe^{++} permits these two ions to be substituted for each other in the crystal, the links between the silicate groups may be occupied with either iron or magnesium, or by a mixture of the two. This arrangement produces a wide variety of olivines, olive green in color.

If the simple silicate groups are connected to each other through an oxygen atom that is mutually shared by two silicon atoms—an oxygen bridge—we may expect to find the silicate groups attached directly to each other in long chains by covalent bonds:



Since the silicon atom is surrounded by four oxygen atoms in a tetrahedral arrangement, this silicate chain would appear as in Fig. 16-5. The unfilled positions on the oxygen

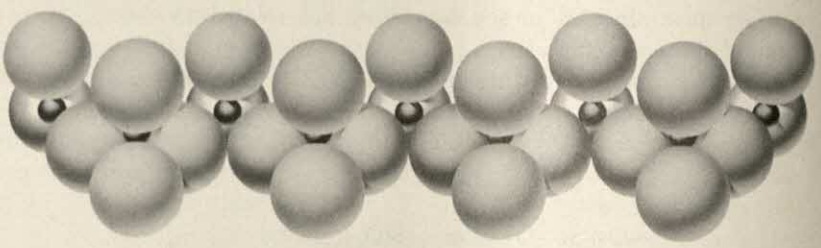


Fig. 16-5. A silicate chain. The small black spheres are silicon atoms and the larger spheres are oxygen atoms. Note that each silicon atom is bonded to four oxygen atoms, and that each oxygen atom is bonded to two silicon atoms or has an unbonded valence that may be bonded to another silicate.

atoms along the side of the silicate chain may be occupied by metal ions, such as magnesium and iron, which in turn may be bonded to another chain of silicate groups. This structure obviously leads into some extremely complex arrangements, and by means of X-ray diffraction analysis, the electron microscope, and other methods we have been able to confirm that these complex structures do exist. The naturally occurring amphiboles, such as asbestos and pyroxenes—for example, the ferromagnesium silicate in granite—are examples of these silicate chain structures.

The most complex silicates, and the most interesting, are those that are bonded in a plane or sheet. If several silicate chains such as those in Fig. 16-5 are attached to each other as they lie side by side, a broad silicate sheet is formed that may be hundreds of times longer and wider than it is thick (Fig. 16-6). These thin crystalline layers supply the pri-

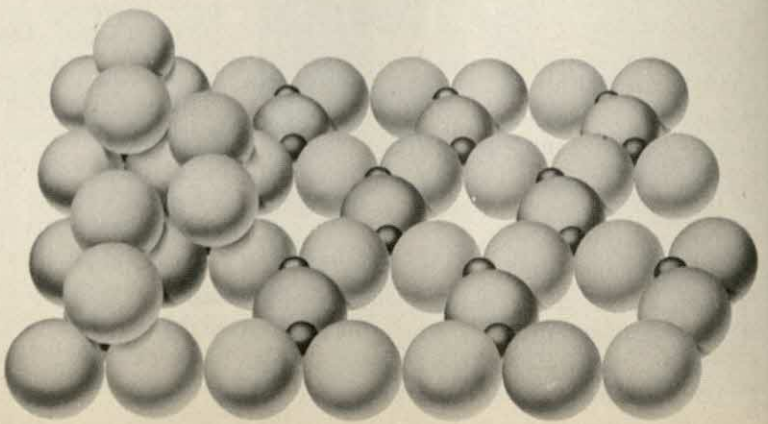


Fig. 16-6. A silicate sheet. Each silicate tetrahedron is surrounded by three others. This structure is found in the micas and in many clays.

mary structure found in clays. In the last chapter we saw how water attacks the components of granite and produces sand, clay, and soluble salts. Now we see that clays, as derived from the feldspar in granite, are complex silicates held together by covalent bonds with definite crystal structures.

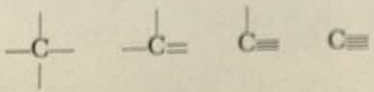
Silicon is not uncommonly found in plant and animal life—in the stems of some grasses and horsetails, in birds' feathers, and in connective tissue, the inelastic fibers that hold together the muscles, joints, and most of the structural parts of our bodies. The silicon in these organisms is in a complex substance often related to similar carbon compounds. We may view this similarity between silicon compounds and carbon compounds as a bridge between organic carbon chemistry and inorganic silicon chemistry. This relationship shows the difficulty in setting up strict classifications without considering the fact that there are usually substances in between these classes that erase any sharp distinction between them.

Ordinary glass is a silicate system consisting primarily of silicon attached by covalent bonds to oxygen and then to certain metals. Window, or lime, glass contains calcium and sodium in addition to silicon and oxygen. Pyrex glass is mostly boron, silicon, and oxygen, with small amounts of sodium and aluminum. Although the bonds in glass are covalent, the atoms in glass are not arranged in crystal form. Instead, when glass cools from a fluid state, it becomes more and more rigid, but retains the lack of order that characterizes liquids. Glass, in short, has a liquid structure, but at ordinary temperatures it becomes too stiff or viscous to flow and retains a rigid structure permanently.

16-4 Structure in Carbon Compounds

The study of carbon compounds is a large undertaking because carbon can form so many bonds both with itself and with other elements. There are certain simple rules, however, that the carbon atom obeys, and a brief review of them will greatly assist us to become acquainted with at least some of the common carbon compounds.

In the first place, carbon almost always has a covalence of 4. That is, in organic compounds the carbon atom is attached through four pairs of shared electrons with other atoms. The arrangement of these four bonds is limited, and if we let a covalent bond be represented by a single dash, we may represent the four possible combinations as follows:



As in the case of silicon, the covalent bonds of the carbon atom are directional, which means that there is a definite relationship between the angles at which these bonds exist.

For example, the first combination of carbon's valence bonds above is a tetrahedral arrangement (Fig. 16-7), but when

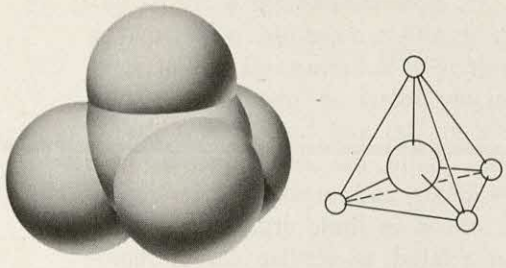


Fig. 16-7. The tetrahedral arrangement of the four covalent bonds of the carbon atom.

two or more of the bonds are attached to another atom the tetrahedral space relationship is distorted, and this distortion produces a variety of space structures in carbon that are unique with this element.

Most carbon compounds contain, in addition to one or more carbon atoms, atoms of hydrogen, the halogens, oxygen, and nitrogen. By investigating the spacing of the covalent bonds of these elements, we can get an idea of how all the other elements are attached in organic compounds. Hydrogen and the halogens have a single covalent bond:

$$H— \qquad Cl— \qquad Br— \qquad F—$$

Oxygen, with a covalence of 2, has two bonds spaced in either of two ways:

$$—O— \qquad O=$$

You will recall that in the water molecule the two hydrogen atoms are attached to the oxygen atom at an angle of 105°. This same angle exists between the two covalent bonds of oxygen in almost all compounds where oxygen is attached to two different atoms.

Nitrogen has a covalence of 3, and the three bonds may be represented in this way:

$$\begin{array}{c}|\\ —N— \end{array} \qquad \begin{array}{c}|\\ N= \end{array} \qquad N\equiv$$

If we construct a simple molecule of nitrogen and hydrogen, we have:

$$\begin{array}{c} H \\ | \\ H—N—H \end{array}$$

This is the ammonia molecule. A careful study of its structure reveals a striking phenomenon: the three hydrogen atoms lie in a plane at the corners of an equilateral triangle, and the nitrogen atom is either above or below the center of this plane (Fig. 16-8). When radar waves (with wave lengths of 1–5 cm) are passed through a balloon filled with ammonia

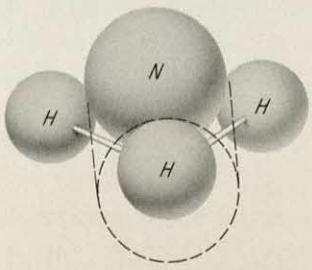


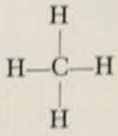
Fig. 16-8. The ammonia molecule.

gas, most of the radar waves pass through unaffected. However, the radar waves of precisely 1.2861 cm wave length are absorbed by the ammonia molecules because the frequency of this particular wave is exactly the frequency of the oscillating nitrogen atom as it jumps back and forth through the plane of the hydrogen atoms. This is an example of interference of wave motion as described in Chapter 10. Since the nitrogen atom moves regularly like the pendulum of a clock, its vibrations can be used to stabilize a radar wave generator, which in turn can check the passage of time. We saw in Chapter 2 that measuring time by the solar day is accurate only to within one second in every 30 million seconds. Our "ammonia clock," on the other hand, can measure time to within one second in ten *billion* seconds. Needless to say, this molecular clock is a big improvement over our astronomical clocks, and we are now able to check closely the accuracy of the solar day and the sidereal day. The fact that the oscillations inside a tiny molecule can be used to measure the sweeping movement of the earth in the solar system indicates that movements within molecules are more precise than movements on the grand scale of the solar system.

Bear in mind the rules of valence the carbon atom obeys as we examine the structure and behavior of the simpler organic compounds.

16-5 Hydrocarbons

A large group of carbon compounds consist of only carbon and hydrogen, and we call these substances *hydrocarbons*. The simplest hydrocarbons are those containing carbon atoms attached to each other in an open chain structure, but let us begin with the hydrocarbon that contains only one carbon atom—methane:



A three-dimensional picture of the methane molecule would show the four hydrogen atoms in a tetrahedral arrangement about the central carbon atom.

Methane is the chief component of natural gas. When used as a fuel it burns with oxygen from the air to form carbon dioxide and water:

$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$$

When methane is burned in a limited amount of air, a form of graphite carbon, called *carbon black*, is produced:

$$CH_4 + O_2 \rightarrow C + 2H_2O$$

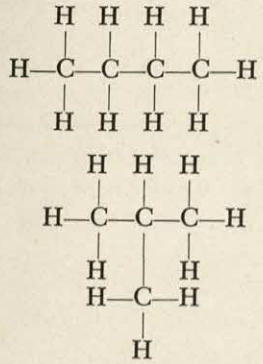
Carbon black is used as a black paint pigment and in processing rubber for automobile tires.

Closely related to methane is a series of hydrocarbons having the general formula C_nH_{2n+2}. The first eight members of this series are:

Methane	CH_4
Ethane	C_2H_6
Propane	C_3H_8
Butane	C_4H_{10}
Pentane	C_5H_{12}
Hexane	C_6H_{14}
Heptane	C_7H_{16}
Octane	C_8H_{18}

Higher members of this series include the components of gasoline (C_7H_{16} to C_9H_{20}) and kerosene ($C_{10}H_{22}$ to $C_{16}H_{34}$). Heavy fuel oils, lubricating oils, petroleum jelly, and solid paraffin contain still larger molecules of hydrocarbons.

The structure of the hydrocarbons makes the space relationships among the atoms in them extremely important. When two or more substances contain the same number of atoms, but the atoms in them are arranged differently, they are called *isomers*. The smallest hydrocarbon in the preceding list that exhibits isomerism is butane, the atoms of which may be arranged in two different structures:



In keeping with the tetrahedral arrangement of the four bonds of the carbon atom, the two butane isomers have the structure shown in Fig. 16-9. The zigzag shape of the carbon chain results from the directional bonding of the carbon atoms. The greater the number of carbon atoms in the hydrocarbon, the greater the number of isomers. Heptane, for example, has 9 isomers; octane has 18 isomers; and the hydrocarbon $C_{40}H_{82}$ theoretically can have almost 63 *trillion* possible isomers.

The origin of petroleum deposits is not definitely known, but it is believed that petroleum and natural gas are the remains of marine sediments that have rotted or putrefied over a long period of time. The floor of the Black Sea may at this time be generating a petroleum deposit, because 35 per cent of the sediments on its bottom is organic matter, over a dozen times that found in normal sediments. When such black, slimy, organic muds are subsequently overlaid by a

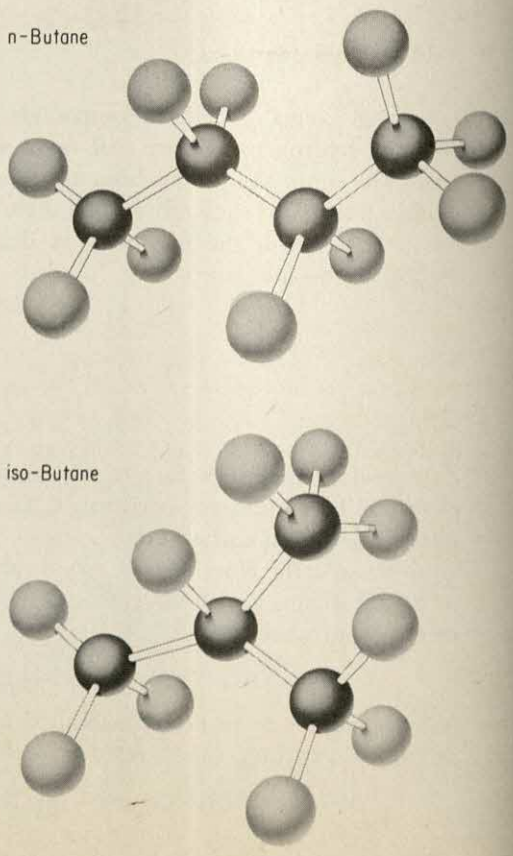


Fig. 16-9. The structures of normal and isobutane isomers.

porous rock such as sandstone, then capped with an impermeable shale, a dome is formed that serves to trap the petroleum. The anticline in Fig. 16-10 is an ideal example of this sort of trap. The shape of a subterranean anticline enables oil prospectors to sound out the petroleum deposit by sending wave motions through the earth's crust (see Chapter 10).

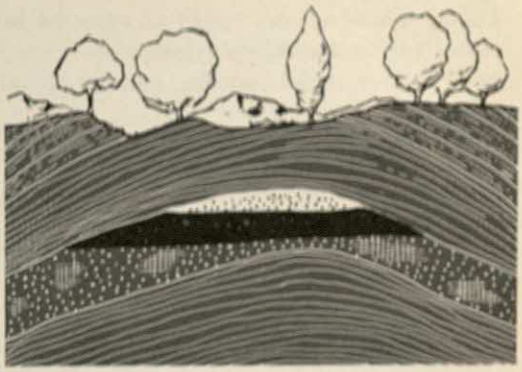


Fig. 16-10. An anticlinal trap for petroleum. Water, being more dense than petroleum, forces the petroleum upwards into the "dome."

Petroleum is divided into its various components by a process known as *distillation*, in which the fact that the components with small molecular weights vaporize at lower temperatures than do the components with high molecular weights is utilized to obtain gasoline, kerosene, gas oil, and lubricating oil. In view of the tremendous demand for gasoline, however, much of the kerosene and gas oil is converted to gasoline by a method called *cracking* (Fig. 16-11). In this process, a typical molecule of kerosene, $C_{16}H_{34}$, which is a long hydrocarbon chain:

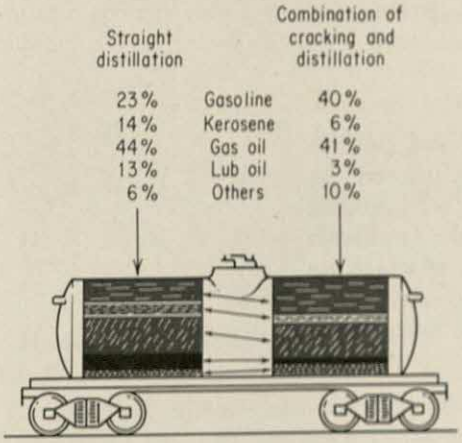
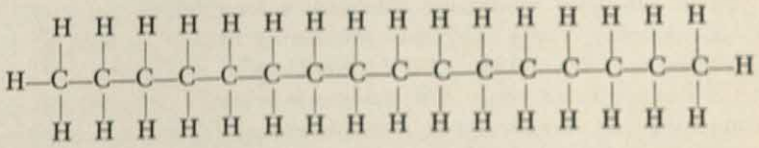
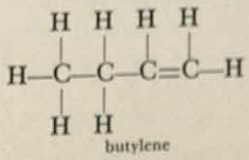


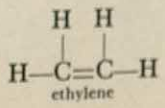
Fig. 16-11. Products from petroleum: (a) from straight distillation, and (b) from a combination of cracking and distillation.



is broken down into smaller fragments at elevated temperatures and pressures. These fragments may include the hydrocarbons hexane, heptane, and octane, and in addition the hydrocarbons such as C_4H_8 (butylene) and C_2H_4 (ethylene) may be split off from the larger molecules. The structural formulas for these last two hydrocarbons show that two of the carbons are bonded with double covalent bonds:

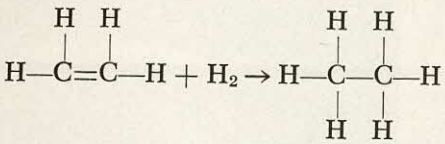


butylene

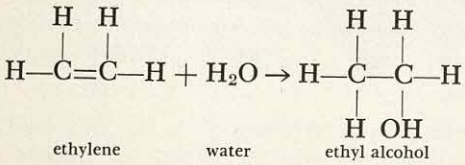


ethylene

These compounds are called *unsaturated* hydrocarbons because they are capable of taking more atoms into their structure. Hydrogen, for instance, may be added to ethylene to give ethane:



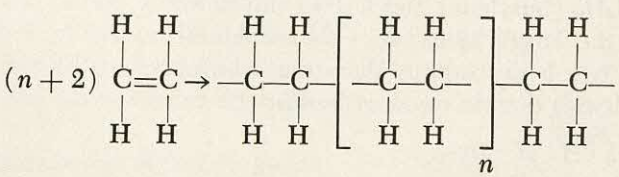
Under suitable conditions, we may add a molecule of water to ethylene and thereby introduce a hydroxyl group into the hydrocarbon molecule:



Such a hydroxylated compound is called an *alcohol,* and this particular one is ethyl alcohol. The petroleum industry, by providing large quantities of cheap ethylene as a by-product of the cracking operation, is responsible for the large-scale production of many common substances, a few of which are ethylene glycol (antifreeze), ethyl ether (an anesthetic), polyethylene (a plastic), mustard gas, and refrigerating agents.

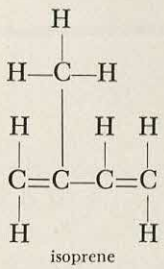
16-6 Polymers

Another type of substance in organic chemistry is one known as a *high polymer.* These substances are similar to the giant molecules we encountered earlier in this chapter, for a polymer is formed of recurring units, each unit being capable of combining with another like it. A simple example of a polymer is the one formed by ethylene combining with itself in which the number n may be several hundred or more:



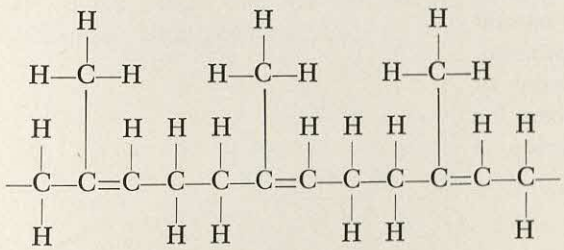
The ethylene molecules continue to combine with each other until long molecular chains are formed. Several thousand ethylene molecules may combine to form a single, huge molecule, and one such substance, called *polyethylene,* is a plastic that is widely used in sheets and in many molded objects; chemically it is similar to paraffin.

Rubber is a high polymer. The basic unit—called *isoprene*—in a natural rubber molecule is a hydrocarbon that is somewhat more complex than ethylene and, since it contains two carbon-carbon double bonds, is readily subject to polymerization:



isoprene

Although natural rubber contains small quantities of other substances, it is made up chiefly of long chains of thousands of isoprene units:



In the process of vulcanization, an atom of sulfur is added to the double bond in each isoprene unit, and since the sulfur atom has a covalence of 2, it unites two adjacent rubber molecules, thus increasing the strength of the rubber polymer. Under normal conditions, the long rubber molecules are twisted and coiled. Rubber is able to stretch because the molecules can uncoil, and, when the rubber snaps back to its original length, the molecules return to their coiled shape.

Synthetic rubber, on the other hand, although similar to natural rubber, is composed of somewhat different units of hydrocarbons that are derived from petroleum cracking. A common form of synthetic rubber is made by polymerizing a mixture of two substances—styrene and butadiene. Styrene is closely related to ethylene, and butadiene, which is quite similar to isoprene, contains two double bonds. Each of these are capable of forming their own polymers; but by polymerizing a mixture of them the polymer chain contains alternately one kind of unit and then the other. This imparts properties to the polymer not possible to achieve with a single component. Such a polymer is called a *copolymer*, and nylon, dacron, and other modern synthetics are copolymers. In a later section, after we become acquainted with the chemistry of the units in these polymers, we will examine these interesting materials in more detail.

A recent improvement in polymers—sometimes viewed as

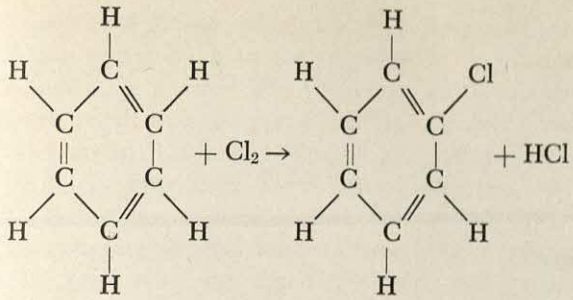
a revolution in this area—has stemmed from our new ability to orient the units to each other as they become part of the polymer chain. For example, imagine a parking lot where each patron is permitted to park his car any place he wishes and contrast that to the case where the attendant carefully parks each car bumper to bumper and all headed the same way. Obviously, in the latter arrangement more cars can be parked in the area and in a more orderly way. Likewise, with polymers, the older methods of bringing the units together resulted in a lack of order, and the resulting polymers were not closely packed and strong. Recently, however, methods of orienting the units "bumper to bumper" in a crystalline fashion have been found, and the resulting polymers are notably stronger, more dense, and in general greatly improved. The structure of natural rubber itself has at last been duplicated by this new method, and the process promises to lead to greatly improved synthetic polymers that can be used as rubbers, fibers, films, and moldings.

In addition to synthetic polymers which have become almost indispensable to modern man, there are many natural polymers such as cellulose, starch, pectins, and proteins. These biological polymers are essential to the foundation of life and intelligence, and we depend on many of them for food. The structure of some of these natural polymers is discussed in Chapter 18.

16-7 Cyclic Hydrocarbons

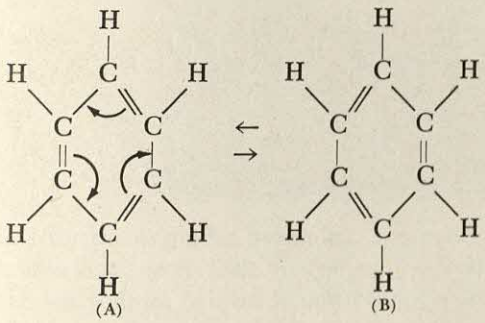
If we heat some soft coal in the absence of air, almost 200 different compounds may be obtained. The most important product of this heating, however, is a liquid that is of great interest in organic chemistry—*benzene*—which is an inflammable, colorless liquid with a rather pleasant odor. A ton of coal yields about 2 gallons of benzene. Benzene baffled the early chemists because from its formula, C_6H_6, they expected it, like ethylene, to have several double covalent bonds. But benzene is contrary, for it does not act as though it contained ordinary double bonds: it does not polymerize and is indeed a rather inert substance. After many years of careful study and much theorizing, chemists developed a picture of the benzene molecule that revealed its inner secret. A formula for the benzene ring was proposed by August Kekule in 1865 that did for organic chemistry what Mendeleev's periodic table did for inorganic chemistry in 1869. Both men opened doors for the next big advance in their respective fields.

Benzene has probably been investigated more thoroughly than any other chemical substance. Each hydrogen atom in benzene is located at a corner of the hexagonal ring, and if we add chlorine to benzene in the presence of ferric chloride, the hydrogen atoms will, one by one, be substituted by chlorine atoms:



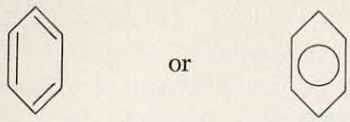
When benzene has two chlorine atoms attached to opposite corners of the ring, it is in the form called *paradichlorobenzene,* or moth crystals. If six chlorine atoms replace all the hydrogen atoms in the ring, we have *hexachlorobenzene,* a powerful insecticide widely used in agriculture.

We have shown three double bonds in the benzene formula in order to account for the double covalent bonds for all six of the carbon atoms. These bonds, however, are not ordinary double bonds; they are believed to be capable of readily changing their positions:



This ability of the bonds to switch between atoms is called *resonance,* which is a modern concept that explains why the benzene molecule is more stable than an ordinary hydrocarbon molecule containing three single and three double carbon-carbon bonds. Both the (A) and (B) structures for benzene are equally probable, and it is assumed that the molecule is a hybrid that lies somewhere between these two structures. Associated with this so-called hybrid structure is a certain amount of energy, called *resonance energy,* which is equal to 198,000 joules per mole of benzene. This is the amount of energy in the molecule in addition to the energy of the regular covalent bonds, and it accounts for the fact that the benzene molecule is more stable than we would expect it to be if it were exactly like either formula (A) or (B) above. The ammonia molecule, considered earlier in this chapter, is a resonating structure though much simpler in arrangement than the benzene resonance.

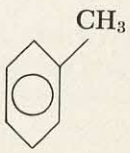
Since we will encounter more compounds that contain the benzene ring, in order to simplify the writing of its formula, we shall indicate benzene by a hexagon:



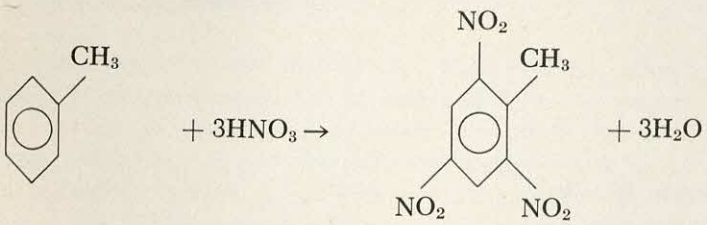
When hydrocarbon chains are substituted for the hydrogen atoms in benzene, a vast array of different compounds results. For instance, to each corner of the benzene ring we may attach hydrocarbon groups that are derived from the simple hydrocarbons. These groups, or radicals, are related to their corresponding hydrocarbons as follows:

Hydrocarbon	*Radical (R)*
CH_4 methane	CH_3—, methyl group
C_2H_6, ethane	C_2H_5—, ethyl group
C_3H_8, propane	C_3H_7—, propyl group

A methyl group, attached to one corner of the benzene ring gives toluene:



One of the most common commercial and military explosives, trinitrotoluene (TNT), is made from toluene and nitric acid:



toluene nitric acid TNT

TNT is an exceptionally effective explosive. It may be handled and transported safely, yet when detonated it releases a tremendous amount of energy. Although the chemical reaction in an explosion of this sort is primarily one of combustion, TNT does not require oxygen from the air in order to explode because it contains sufficient oxygen in itself to cause an explosion. Since the combustion products are all gases such as carbon dioxide, water vapor, and nitrogen, the detonation of an explosive is a rapid conversion of a solid substance into a gas mixture, accompanied by large mechanical forces of expansion.

16-8 Alcohols, Acids, and Esters

The study of organic chemistry is to a great extent a study

of *functional groups*, which are those groups of atoms in an organic compound that are responsible for the compound's distinctive properties. The properties of an alcohol, for example, result in large part from the fact that it contains an —OH group. Table 16-1 lists the more important functional groups. To appreciate the following discussion, it is advisable to become acquainted with these groups, and, although there are several other kinds of functional groups among organic substances, the list in the table will suffice for our present interest.

Alcohols are organic substances that contain one or more hydroxyl (OH) groups. Note that the formula of a simple alcohol, such as ethyl alcohol, is much like that of a base:

C_2H_5—OH	ethyl alcohol
Na—OH	sodium hydroxide

TABLE 16-1 **FUNCTIONAL GROUPS IN ORGANIC COMPOUNDS**
R is any hydrocarbon radical

Type	*Functional group*	*Example* (R equals CH_3—)	
Hydrocarbon	R—H	CH_3—H	methane
Alcohol	R—OH	CH_3—OH	methyl alcohol
Acid	R—COOH	CH_3—COOH	acetic acid
Ester	R—COO—R′	CH_3—COO—CH_3	methyl acetate
Amine	R—NH_2	CH_3—NH_2	methylamine
Amide	R—CO—NH—R′	CH_3—CO—NH—C_2H_5	ethyl acetamide

Beyond their similar formulas, however, these two substances have very little in common. Alcohols are covalent, whereas sodium hydroxide and other typical bases are ionic substances. It is more proper to compare alcohols to water because alcohol, like water, is essentially a covalent substance:

CH_3—OH	methyl alcohol
H—OH	water

Methyl alcohol is sometimes called *wood alcohol* because it may be obtained through the destructive distillation of wood, but it can now be made more cheaply from coal. The coal is first converted to coke, then steam is passed over the hot coke, which, being composed chiefly of carbon, reacts with the steam to form "water gas," a mixture of hydrogen and carbon monoxide:

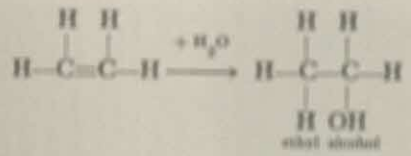
$$C + H_2O \rightarrow CO + H_2$$

The water gas is then passed over a mixture of zinc and chromium oxides at high pressure and at a temperature of 400°C. The metallic oxides serve as a *catalyst*, which is a substance that influences the rate of a chemical change and is recovered unchanged after the reaction has occurred. The water gas is converted into methyl alcohol in the presence of this catalyst:

$$\underset{\text{carbon monoxide}}{CO} + \underset{\text{hydrogen}}{2H_2} \rightarrow \underset{\text{methyl alcohol}}{CH_3OH}$$

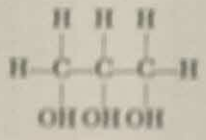
This process is an example of synthetic organic chemistry—that is, the production of an organic substance from simple raw materials—and it is also an example of how scientific knowledge often eliminates our dependence upon a natural source for a useful material. Several synthetics, such as methyl alcohol, synthetic rubber, nylon, and others, have revolutionized the industries in which they are used, and it is little wonder that modern industry spends huge amounts of money to learn more about the products upon which it is dependent.

Ethyl alcohol is made by two processes: synthesis and fermentation. In synthesizing ethyl alcohol, ethylene, C_2H_4—a by-product of the cracking process—may be converted into ethyl alcohol by the direct addition of water:



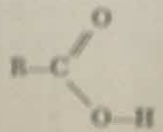
In the fermentation method—a process known since early times—the alcohol is produced by the action of yeast upon sugars. We shall give more consideration to the fermentation process in Chapter 18.

Another common alcohol, one that contains three hydroxyl groups, is glycerol (glycerine), a sweet-tasting, viscous liquid used in cosmetics, food products, and in organic synthesis:



Glycerol is obtained as a by-product in soap manufacture and occurs naturally in vegetable and animal fats, in which it is combined with organic acids.

Although *organic acids* are acids in the ordinary sense—that is, they ionize in water solutions to give hydrogen ions—they are usually quite weak. A partial list of organic acids is given in Table 16-2. Many of these acids are present in foods and they often impart to the food a distinctive flavor, usually a sour taste. The functional group in an organic acid, the carboxyl group, has the structure:



An *ester* is the combination of an alcohol and an acid. In order to form an ester by a reaction between two covalent substances, such as an alcohol and an acid, a third substance

TABLE 16-2 ORGANIC ACIDS

	Formula	Natural occurrence or source
Formic acid	H—COOH	insects (ants)
Acetic acid	CH_3—COOH	vinegar
Butyric acid	C_3H_7—COOH	rancid butter
Stearic acid	$C_{17}H_{35}$—COOH	tallow
Oleic acid	$C_{17}H_{33}$—COOH	cottonseed oil

is usually necessary. A mixture of ethyl alcohol and acetic acid, in the presence of sulfuric acid, will react quite rapidly to form the ester, ethyl acetate:

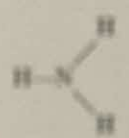
$$\underset{\text{acetic acid}}{CH_3CO\boxed{OH + H}}\underset{\text{ethyl alcohol}}{OC_2H_5} \xrightarrow{H_2SO_4} \underset{\text{ethyl acetate}}{CH_3COOC_2H_5} + \underset{\text{water}}{H_2O}$$

The box in this equation shows that the acid gives up a hydroxyl group and the alcohol loses a hydrogen atom. The elements in the box combine to form water and thus enable the acid and the alcohol to come together to form an ester. The sulfuric acid assists the reaction by combining with the water that is released; in effect it pulls the water away from the alcohol-acid mixture and is therefore a dehydrating agent in this reaction.

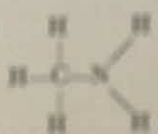
Esters possess distinctive, and often pleasant, odors that are frequently produced by flowers, and such esters may be extracted from the flower, or produced synthetically from suitable alcohols and acids, to provide the basic ingredients of perfumes.

16-9 Amines and Amides

An *amine* is an organic compound whose functional group is related to ammonia. If we write the structure of ammonia:



and replace one of the hydrogen atoms with a methyl group, we have methylamine:



When all three hydrogen atoms in ammonia are replaced with methyl groups we have trimethylamine. All these compounds are basic when in aqueous solutions, and in this re-

spect they are quite similar to ammonia. Amines are often formed in the decomposition of animal tissue, and their odor is a "fishy" one, which is an indication of their organic composition.

Because amines are basic they may be expected to react with organic acids to form an organic salt comparable to the acid-base-salt reaction that we encountered in the last chapter. Here, however, the combination of an amine with an organic acid is called an *amide* instead of a salt. This reaction is similar to the interaction of an acid and an alcohol where an ester is formed. That is, a water molecule is split out between the amine and acid, as shown by the box in the equation:

$$\underset{\text{acetic acid}}{CH_3CO}\boxed{OH + H}\underset{\text{methylamine}}{NHCH_3} \rightarrow \underset{\text{methylacetamide}}{CH_3—CO—NH—CH_3} + \underset{\text{water}}{H_2O}$$

The functional group in an amide,—CO—NH—, is very important in the formation of synthetic polymers between an acid and an amine, such as nylon, and the amide group is also essential to the primary unit that makes up animal tissue (protein).

In making nylon, an acid that contains two carboxyl groups and an amine that contains two amine groups are attached to each other, giving long chainlike molecules. Several different acids and amines may be used, although one of the more popular nylons is composed of adipic acid and hexamethylenediamine. Here is the equation showing how these two react:

$$\underset{\text{adipic acid}}{HOOC—C_4H_8—COOH} + \underset{\text{hexamethylenediamine}}{H_2N—C_6H_{12}—NH_2} \rightarrow$$
$$HOOC—C_4H_8—\underbrace{CO—NH}_{\text{amide linkage}}—C_6H_{12}—NH_2 + H_2O$$

In this equation only one molecule of acid is linked to one molecule of amine, but you can readily see that there is yet an unused carboxyl group in the acid and an unused amine group in the amine, both of which can further react with either another amine or acid molecule to give a long chain polymer. As a result, a single nylon molecule may contain over a hundred molecules of an acid and a like number of amine molecules attached to each other through a series of amide linkages. Since the mechanical strength of nylon is very high, we can conclude that the amide linkage is quite strong. In a later chapter we shall see the importance of the amide linkage in the structure of proteins.

16-10 Summary

Nonmetallic elements, such as hydrogen, oxygen, sulfur, and carbon, are covalent substances, the properties of which are largely related to the covalent structure of their molecular units.

Graphite and diamond are both composed of carbon, but

their radically different properties stem from their different crystal structures. The extreme hardness of diamond results from the fact that its small atoms are held together by short covalent bonds and arranged in a very rigid crystal order whereas graphite is composed of thin sheets.

Organic substances are carbon compounds and inorganic substances are noncarbon compounds.

Silicon forms tetrahedral bonds with oxygen and is capable of entering into both simple and complex substances ranging from quartz to clays and glass.

Carbon also forms tetrahedral covalent bonds and is notably capable of forming a great number of different kinds of molecules. The more important groups of carbon compounds include the hydrocarbons, alcohols, acids, esters, amines, and amides. Carbon, too, forms covalent bonds with itself.

Resonance, as shown by the benzene and ammonia molecules, is a concept that explains the stability and structure of molecules that have two or more alternate structures.

Polymers are giant molecules made up of simple, repetitive units. They occur naturally in rubber, quartz, and the complex silicates, and are produced in a wide variety of synthetics such as synthetic rubber, nylon, and plastics.

EXERCISES

A. KEY TERMS TO REMEMBER

acid (organic)	ester	isomer
alcohol	functional group	organic compounds
amide	giant molecule	polymer
amine	graphite	quartz
ammonia "clock"	high polymers	resonance
benzene	hydrocarbon	tetrahedral arrangement
diamond	inorganic compounds	

B. QUESTIONS ABOUT COVALENT SUBSTANCES

1. Write electronic formulas to show the covalent bond in the molecules: F_2, Cl_2, Br_2, I_2.

2. Explain the hardness of diamond in terms of its crystal structure.

3. With its diamondlike structure, why is tin only a dust at low temperatures? Would it be advisable to use tin pipes in a refrigerator?

4. Starting with a diagram of a silicate chain, as in Fig. 16-5, show how two silicate chains may be attached to each other by magnesium ions. What kind of metal ions would best terminate the chains?

5. Prepare a list of household items that contain silica in one form or another.

6. Draw structural formulas for the isomers of pentane, C_5H_{12}.

7. Assuming that the regular rules of covalency apply to each of the following substances, draw structural formulas for:

a. CH_2O	d. $C_2H_4O_2$	g. C_6H_6	j. C_6H_7N
b. C_2H_6O	e. $C_3H_6O_2$	h. C_7H_8	k. $C_7H_6O_2$
c. CH_2O_2	f. CH_5N	i. C_6H_6O	l. C_6H_{12}

8. Name as many of the substances in the preceding exercise as you can.

9. Prepare a list of household items that are made of synthetic polymers. Trace as many of these as you can to the raw materials from which they are made.

10. How many seconds would the "ammonia clock" gain or lose in 300 years?

11. In view of its crystal structure explain why graphite is a very poor electrical conductor in the direction perpendicular to the plane of the crystal layer.

12. The ring of six carbon atoms in a benzene molecule is thought to be an excellent conductor of electricity. That is, there would be very little resistance to an electrical current if it were induced to flow around the ring. Explain.

13. Construct diagrams that show the electronic formulas of the following covalent substances:

a. S_8, an eight-membered ring of sulfur atoms.
b. P_4, with each phosphorous atom at the corners of a tetrahedron.
c. N_2, involving three pairs of shared electrons.
d. AsH_3
e. Graphite, showing a portion of the layer crystal.

14. What evidence is there that glass is not a true crystalline substance?

15. Explain how the ammonia molecule which is a neutral molecule can form chemical bonds with a molecule of another substance.

16. Using the equation for the partial combustion of methane to carbon and water as a guide, write an equation describing the formation of carbon when a cold surface such as glass or china is placed in a candle flame. Candle paraffin is a mixture, but you may assume the hydrocarbon $C_{20}H_{42}$ to be representative of its composition.

17. Why should a molecule that has a resonance structure be more stable than a similar one without this kind of structure?

18. Write an equation which represents the self-combustion of TNT.

19. Write an equation to show the cracking of $C_{16}H_{34}$ into two smaller molecules. Show the structures of the products and explain why one of the products must be unsaturated.

Seventeen

Chemical Reaction

In previous chapters we have become acquainted with the composition of various simple and compound substances, and we now turn to a key aspect of these substances—how they react with each other in chemical changes. We know that iron tends to rust but gold remains bright and lustrous for years. We also know that wood burns, yet we build houses of it. Many have dreamed of finding a magic pill which, when dissolved in water, would convert the water to a fuel for an engine, and there have been numerous attempts to design an engine that would operate on a fuel that it itself manufactures. But we know that water, sand, concrete, and the various rocks are so little inclined to serve as fuels that these substances are handy to have around as a means of preventing or controlling undesirable combustion reactions. Upon closer examination, it becomes apparent that all chemical reactions involve changes in energy, and the greater this change in energy the greater the tendency for a chemical reaction to occur.

Certain reactions take place very rapidly and with little coaxing on our part. An acid, for example, neutralizes a base almost instantly, and most of the reactions involving ionic substances are rapid reactions and require little of our attention other than mixing the reactants together. Other reactions, however, such as the digestion of food and the disintegration of rocks by natural waters, are slow reactions. Wood must be kindled or heated to its ignition temperature before it burns, and iron rusts slowly.

In this chapter we will be concerned with two major problems: why does a chemical reaction take place, and what determines how fast a reaction goes? We will find that the

answers to these problems involve a serious consideration of the energy changes that accompany chemical reactions. At times we may be more interested in the energy released in a reaction than in the products of the reaction. Nevertheless, in order to obtain this energy, whether it drives our cars, flies our jet planes, or lights a flashlight, chemical reactions occur and the mechanism of these changes supplies a fundamental understanding of why the reaction takes place. Let us seek the relationship between chemical reactions and energy.

17-1 Kinds of Energy

Up to this point we have encountered several types of energy which we defined as the ability to do work. *Mechanical energy* (kinetic and potential) is measured in terms of force and distance (Chapter 2) and is probably the one we are most familiar with in our everyday lives.

Radiant energy is the energy that is transmitted as an electromagnetic wave. We commonly associate the energy of light, gamma rays, X rays, infrared rays, radar, and radio with this form of energy (Fig. 10-27). Radiant energy is composed of quanta or photons, and, according to the quantum theory (Chapter 12), the energy of a photon is directly proportional to the frequency of the radiation. Thus, a photon of ultraviolet radiation has much more energy than does a photon in the infrared portion of the electromagnetic spectrum.

A third kind of energy is *heat energy*. It is measured in terms of a calorie, which is the amount of heat required to raise the temperature of one gram of water from 14.5°C to 15.5°C. The concepts of heat energy are so important in physical science that two later chapters will be devoted to a more complete discussion of it. Heat energy—like mechanical energy and radiant energy—is capable of doing work.

A fourth kind of energy is *electrical energy*. Although we became acquainted with electrical energy in Chapter 9, it is often associated with chemical processes, and it, like heat energy, is important enough to warrant a more thorough discussion later in the book.

The next energy we encounter, and the subject of this chapter, is *chemical energy*. Chemical energy is a sort of potential energy—not the energy of position as is potential mechanical energy, but the energy of electrons in atoms and molecules. A pound of dynamite, for instance, contains a great deal of chemical energy stored up in its molecules, and when the dynamite is exploded, the stored-up chemical energy is released or transformed into other forms of energy. Besides chemical energy, mechanical energy is, of course, produced by the explosion of the dynamite, and heat energy and radiant energy are also involved. Another example of chemical energy is to be seen in the operation of an electrical

dry cell. When we press the switch of a flashlight, we start a chemical reaction that liberates stored-up chemical energy in the form of electrical energy. Chemical processes that produce energy are all about us and, indeed, inside us. The burning of gasoline in an engine, the digestion of food, the heating of our homes with fuels are all examples of chemical energy being released in the form of heat, light, mechanical, or radiant energy.

All these forms of energy can be related to each other through conversion factors whose actual values are not as important as the idea that all forms of energy are equivalent to each other. For example, a calorie of heat energy is equivalent to 3.087 foot-pounds or 4.186 joules, and a calorie of heat energy is equivalent to 0.00116 watt hours of electrical energy. Chemical energy is frequently measured in terms of calories or joules.

17-2 Exothermal and Endothermal Reactions

All chemical reactions involve energy changes, but some give off energy while others use up energy. Let us analyze more closely the important difference between these two kinds of reactions. When we burn coal in a furnace, we are primarily interested in producing heat energy. A pound of high quality coal may release as much as 3,500,000 calories of heat, or 5.5 horsepower-hours. The rusting of iron is a reaction that also gives off heat, though this process is so slow we would need a sensitive instrument to measure the amount of energy released. These two reactions are called *exothermal* reactions because they give off energy when they occur.

It is sometimes convenient to express these reactions in a regular chemical equation. The burning of coal (or carbon) may be described as follows:

$$C + O_2 \rightarrow CO_2 + 94{,}052 \text{ cal}$$

This expression says that one gram atom (12 g) of carbon, when burned to carbon dioxide, gives off 94,052 cal of heat, and it is called a thermochemical equation because it shows the energy transformation as well as the chemical change. There is no way to predict the amount of energy that will be released in such a reaction—it must be measured by experiment.

A reaction, on the other hand, that uses up energy when it takes place is called an *endothermal* reaction. In the preparation of aluminum metal from its ore, Al_2O_3, a tremendous amount of energy must be supplied in order to make the reaction occur:

$$2Al_2O_3 + 815{,}900 \text{ cal} \rightarrow 4Al + 3O_2$$

To obtain the aluminum from two moles of aluminum oxide (204 g), we must supply 815,900 cal of heat or an equivalent

amount of some other form of energy. In an aluminum reduction plant, this energy is usually supplied by electrical energy, and to make the four gram atoms of aluminum (108 g), about 0.9 kilowatt-hour of electrical energy is required, assuming that the operation is 100 per cent efficient. Since the process is only 30 per cent efficient, it is little wonder that aluminum reduction plants must be located near large sources of electrical power.

17-3 The Mechanism of Chemical Reactions

The question at the crux of this section is: "How do two substances react with each other?" You may be surprised to learn that the answer to this question is not fully known, and that many chemical changes are intricate puzzles which the scientist has not yet resolved. He is finding out new things all the time, however. Let us see now what we do know about reactions.

Reactions between ionic substances occur almost instantly. The reason for this lies in the fact that an ion is a charged atom or group of atoms and that the electrical field extends, because of the ion's charge, uniformly in all directions. Suppose we mix a solution of sodium sulfate with a solution of calcium chloride. A white precipitate of calcium sulfate occurs immediately:

$$Na_2SO_4 + CaCl_2 \rightarrow CaSO_4\downarrow + 2NaCl$$

Since ionic substances are involved here, the equation may be more correctly shown:

$$Ca^{++} + SO_4^{--} \rightarrow CaSO_4\downarrow$$

The only ions that take part in the reaction are the calcium ion and the sulfate ion, so when we combine the solutions of sodium sulfate and calcium chloride we are, in effect, merely bringing these two ions into contact with each other. Their opposite electrical charge produces a mutual attraction between these ions, with the result that they interact with each other immediately. Generally speaking, if in a mixture there is a combination of ions that may form an insoluble substance, a precipitate of this substance will occur immediately. In this case, when the calcium ions and the sulfate ions are mixed they form calcium sulfate, an insoluble salt, and the reaction is very rapid because it involves oppositely charged ions that attract each other strongly and quickly. The reaction, caused by the coulombic forces between the ions, is so rapid, in fact, that it is impossible to measure the time interval during which the calcium and sulfate ions react with each other.

Not all reactions between ions produce insoluble precipitates; some produce a gaseous product. For example, the re-

action between baking soda and an acid produces carbon dioxide:

$$\underset{\substack{\text{baking}\\\text{soda}}}{NaHCO_3} + \underset{\substack{\text{acetic}\\\text{acid}}}{HOOCCH_3} \rightarrow \underset{\substack{\text{sodium}\\\text{acetate}}}{NaOOCCH_3} + \underset{\substack{\text{carbonic}\\\text{acid}}}{H_2CO_3}$$

Carbonic acid is an unstable substance that immediately forms carbon dioxide and water:

$$H_2CO_3 \rightarrow H_2O + CO_2$$

If we rewrite this equation in terms of the ionic substances present, we have:

$$Na^+ + HCO_3^- + H^+ + OOCCH_3^- \rightarrow Na^+ + OOCCH_3^- + H_2CO_3$$

By simplifying the equation so that only the ions that react are included, we get:

$$HCO_3^- + H^+ \rightarrow H_2CO_3$$

Here the coulombic forces between two ions of opposite charge bring the ions together to form a new substance, carbonic acid, which in turn decomposes into carbon dioxide and water. Because of the limited solubility of carbon dioxide, it is evolved as a gas. This reaction results whenever baking soda and an acid are used in baking. The acid may be supplied by vinegar (acetic acid), by sour milk (lactic acid), or by lemon juice (citric acid), and the carbon dioxide that is evolved forms tiny gas bubbles in the pastry that causes it "to rise."

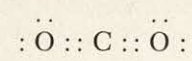
A third kind of product that may be formed in an ionic reaction is a slightly ionized or un-ionized substance. The best example of such a reaction is that between an acid and a base which, as we saw in Chapter 15, is essentially the reaction between a hydrogen ion and a hydroxyl ion:

$$H^+ + OH^- \rightarrow H_2O$$

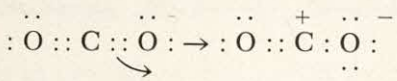
Since water is an un-ionized substance, once the oppositely charged hydrogen and hydroxyl ions are attracted to each other they react rapidly to form water. To sum up, in ionic reactions the driving force is coulombic attraction between ions, and this attraction causes the formation of one of three kinds of products: an insoluble solid, a gas, or an un-ionized substance.

In contrast to ionic reactions, covalent substances usually react slowly. To understand the more complex mechanism of the reaction between covalent substances, let us examine the reaction between two such substances, water and carbon dioxide. We have already seen how a covalent molecule may be polar and thus have some ionic characteristics. Many of the properties of water, for instance, can be explained only by ascribing a polar structure to it (Chapter 15). As we

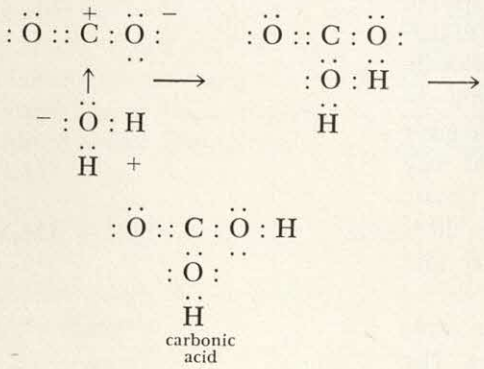
will now see, carbon dioxide is also a polar molecule. The three atoms in carbon dioxide form a linear pattern:



The slightest disturbance of the electron structure of this molecule, however, causes the electrons to vibrate within the molecule and to move first toward one end and then toward the other. At a particular instant, a pair of electrons may shift to the right end:



This change makes the carbon atom positively charged and the oxygen atom on the right negatively charged. If, at this moment, a water molecule happens to collide with the carbon dioxide molecule in such a manner that their polarities are favorably oriented, a molecule of carbonic acid is formed:



The force of attraction between the unlike poles of water and carbon dioxide bring these two molecules together. Note that two steps are involved in this process—the shift of a pair of electrons in the carbon dioxide molecule and a favorable collision with a water molecule. Since the probability of these favorable conditions happening is not high, the reactions of this type take place slowly.

Other reactions are explained by a principle known as the *free radical theory.* It says that when a molecule absorbs a certain amount of energy, in the form of either heat or radiant energy, it can enter into a chemical reaction because the molecule will break in two and will form fragments or radicals that possess an odd number of electrons. The reaction between methane and chlorine provides an example of this mechanism. A mixture of chlorine and methane may be stored in the dark for a long time and not show any evidence of reaction. However, when this mixture is exposed to sunlight a reaction occurs at a rate that is proportional to the intensity of the light. The chlorine molecule first absorbs a quantum of light energy:

$$: \ddot{\underset{\cdot\cdot}{\mathrm{Cl}}} : \ddot{\underset{\cdot\cdot}{\mathrm{Cl}}} : + \text{photon} \rightarrow 2 : \ddot{\underset{\cdot\cdot}{\mathrm{Cl}}} \cdot$$

free radical
(two free chlorine atoms)

The chlorine atom then reacts with a molecule of methane:

$$: \ddot{\underset{\cdot\cdot}{\mathrm{Cl}}} \cdot + \mathrm{H} : \overset{\mathrm{H}}{\underset{\mathrm{H}}{\ddot{\mathrm{C}}}} : \mathrm{H} \rightarrow \mathrm{H} : \ddot{\underset{\cdot\cdot}{\mathrm{Cl}}} : + \mathrm{H} : \overset{\mathrm{H}}{\underset{\mathrm{H}}{\ddot{\mathrm{C}}}} \cdot$$

free radical

The free methyl radical goes on to unite with a chlorine molecule to produce a new chlorine atom and a molecule of methyl chloride:

$$\mathrm{H} : \overset{\mathrm{H}}{\underset{\mathrm{H}}{\ddot{\mathrm{C}}}} \cdot + : \ddot{\underset{\cdot\cdot}{\mathrm{Cl}}} : \ddot{\underset{\cdot\cdot}{\mathrm{Cl}}} : \rightarrow \mathrm{H} : \overset{\mathrm{H}}{\underset{\mathrm{H}}{\ddot{\mathrm{C}}}} : \ddot{\underset{\cdot\cdot}{\mathrm{Cl}}} : + : \ddot{\underset{\cdot\cdot}{\mathrm{Cl}}} \cdot$$

The new chlorine atom will then carry on this process—called a *chain reaction*—until all the free radicals combine with one another, with the walls of the container, or with the reactants. This type of reaction, which involves radiant energy, is called a *photochemical process*, and we will have an opportunity to become better acquainted with it in the next chaper on photosynthesis.

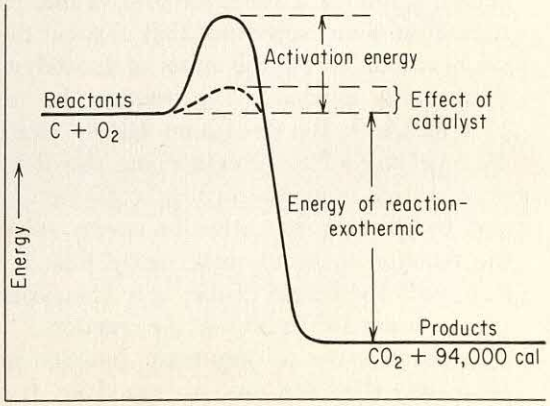
Although there are other ways in which reactions may occur, they are mostly of too special or narrow an interest to be included in this general study. It is sufficient to note here that reactions between ions are usually very rapid, that most reactions involving covalent bonds and dipolar molecules are measurably slow, and that special kinds of reactions, such as photochemical reactions, may be either rapid or slow, depending upon the nature of the reactants.

17-4 Activation Energy and Catalysts

Have you ever wondered why a match does not suddenly and spontaneously burst into flame? Or a log, or a piece of paper, or even a wooden house? These combustion reactions are all exothermic which means that once started, they will take place rapidly, and sometimes disastrously. Fortunately for us, most of these reactions do not occur spontaneously without the addition of some outside energy. We must strike the match, kindle the log, or set fire to the house before they will burn. This energy required to get a reaction started is called the *activation energy*. Sometimes the activation energy needed is quite great, as we find when we try to burn a log, and sometimes it is quite small, as we quickly see when we strike a match. In both cases, however, we must supply energy to get the reactions going.

The diagram in Fig. 17-1 will help us understand the concept of activation energy. Let us examine the reaction in which carbon and oxygen combine to form carbon dioxide.

Fig. 17-1. Diagram of an exothermal reaction showing the activation energy.



At ordinary room temperature, these two reactants may be in contact with each other for a long time and nothing will happen. If we want them to react, we must supply some energy to them, and we usually accomplish this by adding some heat energy to raise their temperature. Enough heat energy must be supplied to elevate the energy of the reactants over the barrier (of activation energy) that prevents them from reacting at lower temperatures. Once the reactants are over the barrier, they will react on their own and give off both the energy of activation and the energy of reaction. Since this is an exothermal reaction, the energy of reaction will supply enough energy to activate new reactants, and the reaction will proceed until the reactants are exhausted.

Examples of activation energy are common in our everyday experience. A spark plug, when sparked, supplies the activation energy needed to set off an explosion of the gasoline-air mixture in an automobile engine. The mechanical energy that we expend in striking a match supplies the activation energy, in the form of heat, necessary to set off the match.

Endothermal reactions also require activation energy, but these reactions are not self-sustaining. An examination of Fig. 17-2 shows that the reactants, nitrogen and oxygen, will react if we supply both the activation energy and enough additional energy to provide the energy that is used up by the reaction itself. Few endothermal reactions require as much activation energy as the one illustrated in the figure, for the reaction between nitrogen and oxygen is a very sluggish one. When the air, which is composed of a mixture of nitrogen and oxygen, is occasionally rent by a bolt of lightning, the tremendous amount of energy supplied by the electrical discharge is sufficient to activate nitrogen and oxygen in a reaction that produces a small amount of nitric oxide. The nitric oxide (NO) then reacts with water and falls to the earth as nitric acid. Small amounts of this acid can usually be found in the vicinity of thunderstorms. At one time, this process was used commercially to make nitric acid, but the electrical energy required proved too costly, and the process was discontinued.

To help overcome the "activation" barrier, a substance called a *catalyst* is often introduced into chemical reactions. A catalyst is any substance that changes the rate of a chemical reaction. To see the effect of a catalyst upon the activation energy needed for a reaction, let us re-examine Fig. 17-1 in which the activation energy has been reduced by the presence of a catalyst (note the dotted curve). Here, the catalyst reduces the height of the "activation" barrier, and, in lowering the activation energy required, thus enables the reaction to occur more easily than it would otherwise. And, with the height of the "activation" barrier lowered, less energy is needed to initiate the reaction.

Catalysts serve an important function in many industrial processes: they are used to speed up the reactions in the

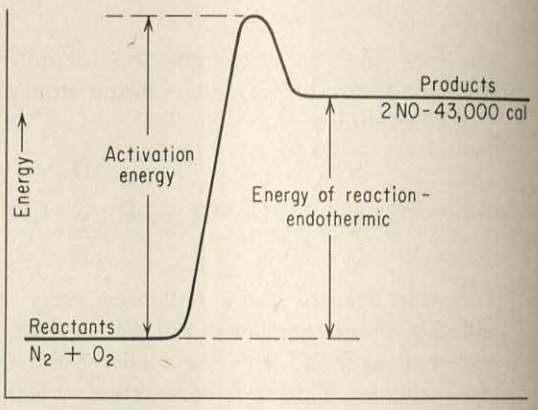


Fig. 17-2. Activation energy in an endothermal reaction.

manufacture of ammonia, sulfuric acid, synthetic gasoline, aviation gasoline, synthetic rubber, hydrogenated oils and fats, and many other useful products. On the other hand, the gasoline we use in our automobiles contains catalysts that elevate the "activation" barrier, thus preventing a premature explosion, or "ping," in the fuel-air mixture.

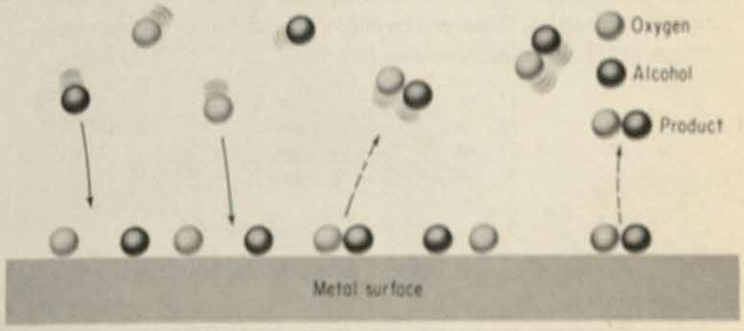
Although many scientists have studied the way in which catalysts change the speed of a chemical reaction, they have as yet failed to develop a theory that satisfactorily explains how catalysts do it. Some catalysts apparently form a kind of intermediate "reaction complex," which works in this way. If two substances A and B react to form the product AB, catalyst C assists this reaction by forming an intermediate substance "AC" or "BC," and this intermediate substance then reacts to form the final product:

$$A + C \rightarrow \text{"}AC\text{"}$$
$$\text{"}AC\text{"} + B \rightarrow AB + C$$

The catalyst C is regenerated and may be used over and over again. Seldom, however, has the intermediate substance "AC" been isolated and examined, and it is usually difficult to prove whether these intermediate substances were ever present in the reaction.

A special breed of catalysts are those called *surface catalysts,* which provide surfaces or areas on which reactions can more readily take place. A platinum wire, for example, if hung in alcohol vapor glows with a red heat, because the surface of the platinum absorbs alcohol molecules and oxygen molecules and thus brings them close enough together to react with each other in an exothermal reaction that is sufficient to make the platinum wire glow (Fig. 17-3). Since the heat of the reaction causes the products to evaporate, the metal surface is left clean and ready to adsorb more alcohol and oxygen molecules.

Fig. 17-3. A probable mechanism for a surface catalyst.



17-5 Rate of Chemical Reaction

As we know, some reactions occur rapidly, and some slowly. In our industrial society, the ability to control the rate of reaction is becoming increasingly desirable, if not, indeed, indispensable. When we make a substance by a

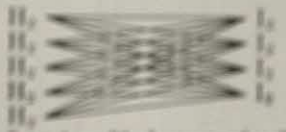
chemical process we would like to have the chemical reactions occur rapidly—we hardly care to wait a week for a pound of oleomargarine to be formed—and, on the other hand, we would like to be able to slow down the oxidation of our rubber tires and our iron bridges. Let us consider the factors that we must understand if we are to control the rate of a chemical reaction.

The first of these factors is the *nature of the reactants*. Iron, for example, rusts slowly but magnesium tarnishes, or rusts, rapidly. Wood burns slowly, but gasoline burns rapidly. To explain these differences, we need to know about the energy changes of all the electrons involved in these processes, and this is a complex story, so complex in fact that scientists at present cannot completely fathom it. All they know is that there are inherent differences among substances that determine how rapidly they will react, and they still have to rely on experiments to find those substances that react rapidly and those that react slowly.

A second factor in controlling the speed of a reaction is the *concentration of the reacting substances*. If a given volume contains twice as many reacting molecules at time *A* as it does at time *B*, at time *A* there will be twice as many chances for a reaction to occur as at time *B*. Stated another way, in what is known as the fundamental *law of mass action, the speed of a reaction is proportional to the concentration of the reactants*. Let us apply this concept to the simple reaction between hydrogen and iodine:

$$H_2 + I_2 \rightarrow 2HI$$

If we have a small box that contains one molecule of H_2 and one molecule of I_2, there is *one* chance for a collision between the two molecules. If we place another H_2 molecule in the box, there are *two* chances for a collision. By adding more H_2 and I_2 molecules, we can see that the number of possible chances for a collision between H_2 and I_2 molecules is the product of the number of H_2 molecules times the number of I_2 molecules. If there are five H_2 and four I_2 molecules, there are twenty chances for a collision between an H_2 and an I_2 molecule:



$5 \times 4 = 20$ chances of collision

The rate of a reaction depends upon the number of collisions between the reacting molecules, but, since a collision does not always result in a reaction, to predict the actual rate of reaction, we must know in what fraction of the collisions the molecules react with each other. In most ionic reactions each collision results in a reaction, but in slow reactions only the small fraction of the collisions that are favorable to reaction will actually result in reaction. The con-

centration of reactants is largely determined by the size of the particles of the reacting substances. For example, a pile of logs is not nearly so easy to burn as a pile of kindling, and, if the wood is ground into sawdust, or wood flour, it may easily undergo combustion explosively. The combustion of such dusts has caused many serious catastrophes.

Another factor that affects the speed of a reaction is the *temperature*. In an earlier section of this chapter we found that the activation energy needed to start a reaction is often supplied by heat, which is essentially a rise in temperature, and the greater the amount of heat added the more rapidly the reaction will take place. After many experiments, scientists have observed that a reaction's rate is usually doubled each time the temperature is raised 10°C.

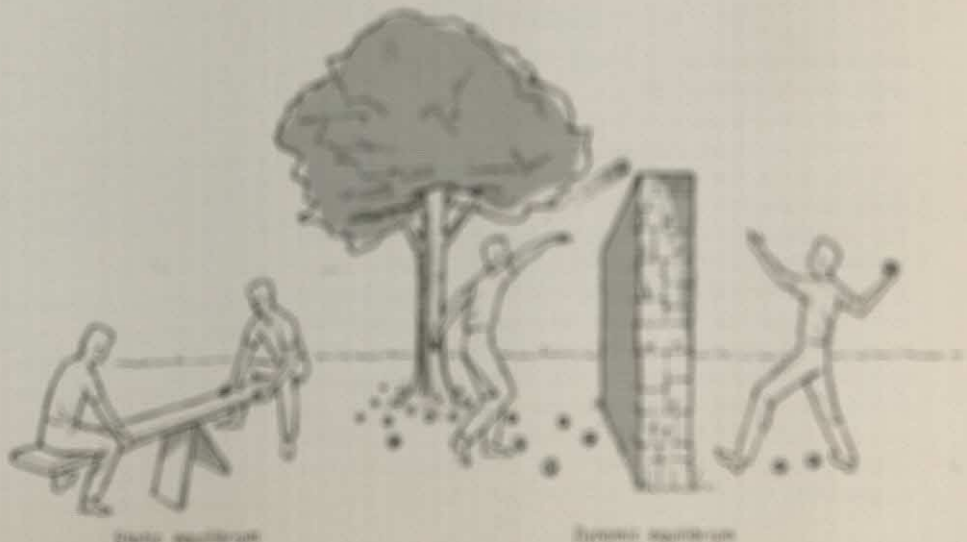
A fourth factor that may alter the speed of a reaction is the presence of a *catalyst*, and we have already examined the role of a catalyst in reaction mechanisms. Since some catalysts speed up reactions and others slow them down or inhibit them, our choice of a catalyst depends on whether we want a particular reaction to proceed faster or slower than it would normally.

17-6 Chemical Equilibrium

As we already know, when two forces or processes act in opposite directions, each exactly counterbalancing the other, the system is in *equilibrium*. When we sit or stand we are in a state of equilibrium. If two boys pull with equal force on opposite ends of a rope, they are in equilibrium. These systems are in the state of equilibrium we call *static equilibrium*, because the system as a whole is not in motion.

But the equilibrium we are concerned with in chemical reactions is another type, called *dynamic equilibrium* (Fig. 17-4). If a boy on the inside of an orchard wall throws apples

Fig. 17-4 Static and dynamic equilibrium.



Static equilibrium

Dynamic equilibrium

at a boy on the outside, the boy on the outside cannot throw any apples back until he can pick up some of the apples thrown over by the first boy. If we sit on the wall and count the number of apples going both ways and find that just as many are going one way as are going the other, we can conclude that a state of dynamic equilibrium exists—one boy, in effect, is undoing the other boy's work.

As in the case of the two boys throwing apples across the wall, many chemical reactions reach such a condition of dynamic equilibrium. For example, if we bring live steam and hot iron together, they will react to produce iron oxide and hydrogen:

$$3Fe + 4H_2O \rightarrow Fe_3O_4 + 4H_2 \qquad (A)$$

But if we keep the hot iron oxide and hydrogen in the same container, they will react and form iron and steam again:

$$Fe_3O_4 + 4H_2 \rightarrow 3Fe + 4H_2O \qquad (B)$$

The second reaction is exactly the opposite of the first one; each reaction undoes the effect of the other, and we speak of this kind of a chemical reaction system as a *reversible reaction.*

At first, when the hot iron and live steam are brought together, only reaction (A) occurs, but as this reaction proceeds, the amounts of iron and steam become smaller and, accordingly, reaction (A) slows down. During reaction (A), increasing amounts of iron oxide and hydrogen are formed until the point is reached at which their amounts become significant enough to react with each other in reaction (B). At the time when reaction (A) has slowed down and reaction (B) has speeded up till they are both operating at the same speed, the reaction system has reached what we call *chemical equilibrium.* Such a condition is indicated by double arrows. These double arrows tell us that this reaction,

$$3Fe + 4H_2O \leftrightarrows Fe_3O_4 + 4H_2 \qquad (C)$$

ends up in equilibrium no matter whether it starts with reaction (A) or reaction (B).

At the point of chemical equilibrium, all substances are not present in equal amounts, for each reversible chemical reaction has its own peculiar equilibrium point, which depends on a number of factors. The forward reaction—reaction (A)—may go almost to completion before the reverse reaction—reaction (B)—starts to take place, in which case the equilibrium point of reaction (C) is said to be displaced far to the right. Or, conversely, the equilibrium point may be displaced to the left. The essential feature in chemical equilibrium is that the forward and reverse reactions operate at the same speeds. Even after the equilibrium point has been reached, both forward and reverse reactions continue to take place, and the system exists in a state of dynamic equilibrium.

Many chemical reactions used in industrial processes—in, for example, the manufacture of sulfuric acid, ammonia, nitric acid, alcohol, steel, and many other products—are reversible reactions and hence obey the principles of dynamic equilibrium. Because the equilibrium point is of great importance in the making of these substances, let us examine the factors that influence it in reversible chemical reactions.

17-7 Factors That Affect Chemical Equilibrium

Since chemical equilibrium consists of two opposing reactions, any change we make that favors one of these reactions will tend to displace the equilibrium point; that is, a change in conditions usually favors one of the reactions more than the other and thus causes a disturbance of the equilibrium. An important guide to predicting the behavior of an equilibrium reaction in these circumstances was formulated in 1888 by a Frenchman, Henri Le Chatelier, and is called the *rule of Le Chatelier* after him. After observing many chemical equilibria, Le Chatelier concluded that *if an equilibrium system is subjected to a change of conditions, the new equilibrium point established as a result of the change tends to undo the effect of the change.* In other words, although a change of conditions favors one reaction in an equilibrium more than it does the other, the equilibrium point will shift in such a direction as to undo the effect of the changed conditions. Let us now examine those changes of conditions that affect a system in equilibrium.

A *change in the concentration* of any of the substances in an equilibrium mixture causes a displacement of the equilibrium. As we saw earlier in this chapter, the rate of a chemical reaction may be influenced by the concentration of the reacting substances, and, consequently, in an equilibrium system a change of the concentration of any of the substances speeds up or slows down one of the reactions and therefore disturbs the equilibrium point. For example, in the reversible reaction between nitrogen and oxygen, an increase in the concentration of oxygen increases the number of chances for the forward reaction and thus speeds up the forward reaction. Since the speed of the forward reaction now exceeds that of the reverse reaction, the equilibrium is disturbed, or shifted to the right. However, the system,

$$N_2 + O_2 \leftrightarrows 2NO$$

overcomes this imbalance, because as the forward reaction produces more product, the concentration of NO becomes large enough to produce a more rapid reverse reaction. Eventually the reverse reaction is able to undo the effect of the forward reaction and a new equilibrium point is established.

A *change in temperature* always affects an equilibrium system. You will recall that all chemical reactions are either

exothermal or endothermal, and, in a reversible chemical reaction, one reaction is exothermal and the other is endothermal. If the forward reaction evolves a certain amount of heat energy, the reverse reaction uses up exactly the same amount of heat energy. If we raise the temperature in such a reaction system, we add heat and this disfavors the forward, or exothermal, reaction. For example, in the following equilibrium reaction:

$$Zn + S \leftrightarrows ZnS + 46{,}000 \text{ cal}$$

each mole of zinc sulfide, when made from zinc and sulfur, releases 46,000 cal of heat energy, but when a mole of zinc sulfide is decomposed into zinc and sulfur, 46,000 cal of energy is used up. Since in the forward reaction, heat is a product of the reaction, as we add heat we discourage the forward reaction and at the same time encourage the reverse reaction. The equilibrium, therefore, is shifted to the left when the temperature is raised.

A *change in pressure* affects an equilibrium system if there is a change in volume as the reaction takes place. In the decomposition of calcium carbonate into calcium oxide and carbon dioxide, a solid is changed into another solid and a gas:

$$CaCO_3 \leftrightarrows CaO + CO_2$$

Since a mole of gas occupies several hundred times the volume of a mole of solid, as this reaction goes forward there is a great increase in volume. If we subject this reaction system to an increased pressure, the forward reaction is disfavored, the reverse reaction is favored, and the original equilibrium is shifted to the left, toward the side that occupies the smaller volume. If a reaction is not accompanied by a change in volume, then a change in pressure has no effect upon the equilibrium point.

The *presence of a catalyst* in a reversible chemical reaction has no effect upon the equilibrium point. Although a catalyst may reduce the time required to reach a state of equilibrium, no catalyst known favors one reaction more than the other in an equilibrium system, so its presence does not affect the relative amounts of reactants and products at equilibrium. If a catalyst were discovered that favored one, but not the other, of the reactions in an equilibrium system, it would be a truly great discovery, because such a catalyst would enable the scientist to conduct many chemical reactions that are now impossible or extremely inefficient.

17-8 The Ammonia Synthesis

In many industrial chemical processes, the proper control of reaction rates and chemical equilibrium makes the difference between success and failure. To see how these reactions are controlled, let us examine a particular industrial

process—the synthesis of ammonia. Prior to 1900, nitrogen compounds were obtained principally from natural sources —coal and nitrates. Then in 1906, the cyanamide process was developed, followed in 1913 by the Haber process (a method of making ammonia from air and water), which today produces most of the world's supply of nitrogen compounds. From the air comes nitrogen and from water comes hydrogen and they are brought together in the following reaction:

$$N_2 + 3H_2 \leftrightarrows 2NH_3 + 22{,}000 \text{ cal}$$

This reaction supplies us with the ammonia that we convert into a multitude of useful substances: fertilizers, nitric acid, explosives, lacquers, plastics such as nylon, and a host of others. It is, indeed, of considerable strategic importance in both war and peace.

When a mixture of one mole of nitrogen and three moles of hydrogen is subjected to various temperatures and pressures, a certain percentage of the mixture is converted into ammonia (Table 17-1). From this table it is quite evident that temperature and pressure have a great effect upon the yield of ammonia, and, ideally, it looks like we should run the reaction at a temperature of 200°C and at a pressure of 1000 atm to get a 99 per cent yield of ammonia. But at temperatures as low as 200°C, we would have to wait some time

TABLE 17-1 PERCENTAGE OF 1:3 NITROGEN-HYDROGEN MIXTURE CONVERTED TO AMMONIA IN THE REACTION $N_2 + 3H_2 \leftrightarrows 2NH_3 + 22{,}000$ CAL

Temperature	*Pressure (in atm)*				
	200	400	600	800	1,000
200°C	90%	96%	97%	98%	99%
300°C	64%	76%	82%	88%	92%
400°C	39%	55%	65%	73%	80%
500°C	20%	32%	43%	51%	57%
600°C	8%	16%	24%	28%	32%
700°C	4%	9%	12%	15%	16%

for the equilibrium point to be reached, and, even if we used the best catalysts available, the stockholders of a company making ammonia under these conditions would be up in arms. Therefore, the reaction is speeded up by increasing its temperature, even though this reduces the percentage yield. Most of the synthetic ammonia plants in the United States operate at a temperature of about 450°C, using an iron-potassium oxide-aluminum oxide catalyst.

From the equation for the ammonia synthesis reaction, we see that four molecules of reactants are converted into two molecules of product, which means that as the two gases react with each other there will be a decrease in volume. An increased pressure, therefore, favors the forward reaction, and this is confirmed in Table 17-1 where we can see that an increase in pressure increases the percentage yield of ammonia and therefore moves the equilibrium point to the

right. For instance, at 700°C, only 4 per cent of the reactants are converted into ammonia at 200 atm pressure, while at 1000 atm the yield is four times greater, for 16 per cent of the reactants are converted into product. Consequently, to produce the largest quantity of ammonia possible, the reaction should operate under a high pressure, but, because of the costly equipment required for such pressures, most of our ammonia is produced at no more than 800 atm pressure. From the table, we can see that at 450°C and 800 atm the maximum yield of ammonia is approximately 65 per cent. A catalyst is always used in the Haber synthesis to reduce the time required to reach equilibrium; it has no effect, however, upon the yield of ammonia at equilibrium.

The concepts of equilibrium and reaction rates are essential ones in most chemical manufacturing, and careful attention to the four factors—temperature, pressure, catalyst, and the nature of the reacting materials—is a prerequisite to successful operation of the chemical industry.

17-9 Reaction Mechanisms in Natural Processes

In Chapter 15, we saw briefly how ground waters attack limestone, feldspars, and other rocks and minerals. These reactions, though sometimes complex, are subject to the same principles that govern the speed of reactions in the laboratory, and we are now in a position to examine some of the factors that control the speed of such natural processes as chemical weathering, sedimentation, and other geochemical changes.

The formation of the earth's crust may be divided into three stages. The first stage, in which all materials were either liquids or gases, occurred at high temperatures, and the less dense substances, such as aluminum, silicon, and oxygen, goes to the surface while the heavier iron and nickel sank to the core of the earth. Because of the high temperatures, these reactions proceeded rapidly, and this stage was of shorter duration than the later stages that operated at much lower temperatures. In the second stage, the earth's surface cooled, and the solid crust that was thereby formed was well above its present temperature. There is ample evidence to show that the hot solid crust of the earth underwent vigorous reactions, some of which no doubt consisted of the reaction between the atmosphere and the earth's crust, and during this second stage most of the elements were converted into their oxides.

The third stage in the development of the earth's crust, now in progress, is operating at a temperature below the boiling point of water. As a consequence, the reactions are much slower than those that took place in the earlier stages, and, in addition, they depend primarily upon the presence of water, carbon dioxide, and oxygen. It is the action of these three upon rocks and minerals at typical crustal temperatures

that is called *weathering*, a process that we shall now examine in the light of what we know about reaction rates and equilibrium.

Natural waters, you will recall, contain a small amount of carbon dioxide. There are two equilibria in the interaction of these two substances because two interdependent reactions are involved:

$$\underset{\text{water}}{H_2O} + \underset{\substack{\text{carbon}\\\text{dioxide}}}{CO_2} \leftrightarrows \underset{\substack{\text{carbonic}\\\text{acid}}}{H_2CO_3} \leftrightarrows \underset{\substack{\text{hydrogen}\\\text{ion}}}{H^+} + \underset{\substack{\text{bicarbonate}\\\text{ion}}}{HCO_3^-} \qquad (D)$$

This equilibrium system contains five substances, and a change in the concentration of any one of them causes both equilibria to shift. For instance, if some of the hydrogen ions are removed from the system, both equilibria will be shifted to the right because removal of hydrogen ions disfavors the reverse reaction. Or, since carbon dioxide is the only gas in this reaction, an increase in pressure favors the reaction that uses up the carbon dioxide and hence produces more carbonic acid and, in turn, more hydrogen and bicarbonate ions. An increase in temperature decreases the solubility of carbon dioxide in water which, in effect, removes some of the carbon dioxide from the environment of the reaction and disfavors the forward reaction. Therefore, the equilibria are shifted to the left at higher temperatures. All these deductions are based on the Le Chatelier rule.

Suppose we investigate what happens when limestone is attacked by water. Limestone is chiefly calcite (calcium carbonate), and in pure water only a small amount of it dissolves, producing a dilute solution of calcium and carbonate ions:

$$CaCO_3 \rightarrow Ca^{++} + CO_3^{--} \qquad (E)$$

However, natural waters in which the calcite dissolves may contain carbon dioxide, and some of the hydrogen ions in the natural water (see Equation D) react with the carbonate ion, forming the bicarbonate ion:

$$\underset{\substack{\text{hydrogen}\\\text{ion}}}{H^+} + \underset{\substack{\text{carbonate}\\\text{ion}}}{CO_3^{--}} \rightarrow \underset{\substack{\text{bicarbonate}\\\text{ion}}}{HCO_3^-} \qquad (F)$$

In contrast to the very slight solubility of calcium carbonate, calcium bicarbonate is quite soluble. Thus, the small amount of carbonate ion supplied in Equation (E) is steadily converted into the bicarbonate ion in Equation (F), with the result that more calcite may dissolve. The weathering of limestone, then, is essentially the slow dissolution of calcite by the carbon dioxide in natural water. We were introduced to this process in Chapter 15, but we can now more fully examine it because it involves several basic ideas of chemical equilibrium.

The speed with which water dissolves calcite is largely determined by the concentration of hydrogen ions in the water. In the laboratory, we can dissolve a piece of limestone in an acid such as hydrochloric acid in only a few

minutes, but in nature the concentration of hydrogen ions supplied by reaction (D) above is only about a millionth of that in hydrochloric acid, so it takes nature a much longer time to dissolve the same piece of limestone. Nevertheless, wherever ground water comes in contact with limestone or other calcium carbonate rocks, it slowly dissolves them.

Although the weathering of feldspars and other silicate rocks is a complex process, the decomposition of feldspar by an acidic system is a rather simple mechanism, called *ion exchange*. The feldspar, $KAlSi_3O_8$, may be converted into clay by the reaction:

$$2KAlSi_3O_8 + 2H^+ + 9H_2O \rightarrow [\text{Intermediate products such as hydrous oxides and cations}] \downarrow Al_2O_3{\cdot}2SiO_2{\cdot}2H_2O + 2K^+ + 4H_4SiO_4 \quad (G)$$

kaolinite clay; silicic acid

In this reaction, the potassium ions in the feldspar are displaced or exchanged by hydrogen ions. Along with this exchange of ions, the feldspar undergoes a major change in crystal structure, but this complex change need not concern us here. It is the presence of hydrogen ions, supplied by natural waters in Equation (D), that makes reaction (G) possible.

The decomposition of feldspar by reaction (G) is also achieved by plants. There are hydrogen ions in the organic material of which the plant's roots are composed, and these hydrogen ions are available for exchange with the potassium ion in the feldspar (Fig. 17-5). The plant requires the potassium ion as a nutrient, and in the growing process the plant obtains this ion, along with other metallic ions, from the feldspar in the soil. A significant amount of clay is made from feldspar by this process. Although the effects of pressure are not obvious, temperature and concentration are critically important in plant growth. We know, for example, that plants can take up soil nutrients much more rapidly in warmer weather, and the farmer is seriously concerned about the concentration of hydrogen ions (the pH) in his soil, because if the soil is too acid the reaction shown in Fig. 17-5 will be impaired.

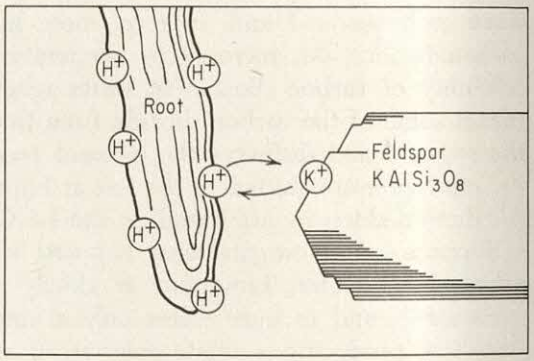


Fig. 17-5. The decomposition of feldspar by plant roots. The hydrogen ions on the rootlet are exchanged with the potassium ion in the feldspar, thus forming clay. The potassium ions are utilized by the plant. Redrawn from Keller and Frederickson, *American Journal of Science*, 250 (1952), 603.

17-10 Summary

In addition to mechanical, radiant, electrical, and heat energy, there is chemical energy—a form of potential energy that is associated with the electrons in the atoms of substances.

All chemical reactions involve a change in chemical energy. When energy is released in a chemical change it is an exothermic process, and when energy is used up it is an endothermic process.

Two substances may react with each other in a variety of

mechanisms. The most common mechanisms include the interaction of ions of opposite charge (rapid reaction), the interaction of dipolar substances (rapid or slow reactions), and the activation of free radicals by either heat or radiant energy (rapid or slow reactions). Reactions tend to go to completion when the product is either a gas, an insoluble substance, or a slightly or nonionized product.

Before a reaction will occur it must be supplied with outside energy called activation energy. A catalyst may either decrease or increase the amount of activation energy needed to make a reaction occur. Catalysts may simply furnish a surface on which the reaction occurs, or they may enter into the process and form an intermediate substance, from which the catalyst is finally recovered.

The rate of a chemical reaction is influenced by the nature and the concentration of the reactants, the temperature, and the pressure. The effect of these factors is governed by Le Chatelier's rule.

Chemical equilibrium is a dynamic balance of the forward and reverse reactions in a reversible chemical process. The equilibrium point may be shifted by a change in concentration, temperature, or pressure.

Natural weathering processes involve the basic concepts of reaction rates and chemical equilibrium. The dissolution of limestone and the conversion of feldspar to clay are two examples of the weathering process.

EXERCISES

A. KEY TERMS TO REMEMBER

activation energy	endothermal reaction	Le Chatelier's rule
catalyst	equilibrium point	reaction rate
chemical energy	exothermal reaction	reversible reaction
chemical equilibrium	free radical theory	static equilibrium
dynamic equilibrium	law of mass action	

B. QUESTIONS ABOUT CHEMICAL REACTIONS

1. What is meant by the expression "mechanical equivalent of heat," and, in general, the "interconvertability of energy"?

2. Classify the following changes as either exothermal or endothermal:

   a. melting of ice
   b. burning of gasoline
   c. neutralization of an acid by a base
   d. rusting of an iron nail
   e. winding a clock

3. Explain why ionic substances, such as acids and bases, react with each other much more rapidly than do covalent substances.

4. What is the role of activation energy in the striking of a match?

5. Is it possible for a catalyst to increase the activation energy of a reaction and, at the same time, slow the reaction down? What is the purpose of the tetraethyl lead in gasoline?

6. What are four factors that may alter the rate of a chemical reaction?

7. Indicate the effects of increased temperature and increased pressure upon the following gaseous equilibria:

$$2SO_2 + O_2 \rightleftarrows 2SO_3 + 45{,}000 \text{ cal}$$
$$H_2 + I_2 \rightleftarrows 2HI + 4{,}000 \text{ cal}$$
$$2H_2 + O_2 \rightleftarrows 2H_2O + 136{,}000 \text{ cal}$$
$$CO_2 + 2SO_3 \rightleftarrows CS_2 + 4O_2 - 265{,}000 \text{ cal}$$

8. Why is a synthetic ammonia plant operated at 450°C when a much greater yield of ammonia could be obtained at lower temperatures?

9. Explain the mechanism by which calcium carbonate is dissolved by natural water.

10. A common sedimentary rock is limestone, which is chiefly calcium carbonate. Explain how limestone, on the basis of reaction (E) on p. 347, may have been deposited from the water in a great inland sea.

11. Catalysts frequently are capable of speeding up certain chemical reactions, but why is it unlikely that a catalyst will be found that can displace an equilibrium point?

C. PROBLEMS ABOUT CHEMICAL REACTIONS

1. How many kilowatt-hours of electricity are needed to make a ton of aluminum, assuming that the process is 100 per cent efficient? How much electricity would actually be needed?

2. How much faster will a reaction occur at a temperature of 40°C than at 10°C? Ans. Eight times faster. Is there any advantage in storing rubber tires, flashlight batteries, and camera film in a cool instead of a warm place? Why?

3. One of the simpler reaction mechanisms is apparently involved when hydrogen and iodine react to form hydrogen iodide: $H_2 + I_2 \rightarrow 2HI$. This involves a favorable collision between H_2 and I_2 and an intermediate substance, (H_2I_2), is supposedly formed. Sketch this favorable collision and show the structure of the intermediate substance and how it breaks up to form two molecules of hydrogen iodide.

4. What would be the maximum percentage yield of ammonia in the Haber process if it were operated at 600 atm and 500°C?

Eighteen

Biophysical Processes

We may not travel far in certain fields of physical science before we cross over into areas of biological science. For instance, the more intimately we become acquainted with organic substances the closer we come to botany, zoology, and bacteriology—the life, sciences. Even those areas of physical science dealing with atomic structure, light energy, and intermolecular forces are productively applied to the major problems in the life sciences. Not content to stop with the discovery of fundamental knowledge within a narrow field, scientists have utilized the discoveries in one field in other, seemingly unrelated fields, and, consequently, some amazing advances have been made. Through these efforts, for instance, our life span has been considerably extended through the control of diseases such as tuberculosis, pneumonia, and poliomyelitis, which are now on the wane; and we can live healthier, happier lives because vitamins essential to our diets have been identified, produced synthetically, and added to our foods when necessary.

In this chapter we will only be able to survey some of these bridge areas between the physical and the biological sciences, but such an introduction should at least open the door enough to provide a glimpse into the fascinating region that lies between these two sciences. In this new region we encounter areas of study called *biophysics*, *biochemistry*, *biogeology*, and other areas so new that scientists themselves have yet to name them. There is general agreement that the

breakthroughs necessary to solve the problems of cancer, heart disease, mental disorders, and old age will probably come with the application of basic, scientific principles to the problem at hand. Thus, we may see a physicist studying the relationship of radiation and cancer, or a chemist making a new complex organic compound with which to combat heart disease, or a physical chemist examining the molecular structure of the genes and chromosomes within the living cells of a mental patient, or a biochemist probing into the chemistry of metabolism.

Even though life itself remains a mystery, certain facets of life processes have been revealed. The basic unit of living systems, the cell, contains substances whose specific purposes have been defined, and even in larger organisms such as man, many of the chemical changes that occur are well enough understood to merit our attention here. In all these processes we should be mindful of the fact that energy is always associated with chemical changes. Plants synthesize many substances such as carbohydrates by using up energy from the sun, and one of the primary needs of our bodies is the energy derived from the foods we eat.

18-1 Autotrophic and Heterotrophic Organisms

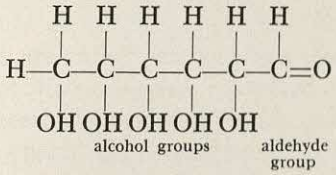
An *autotrophic organism* is a living unit capable of nourishing itself from the simple materials in its environment, and a *heterotrophic organism* is one that depends upon organic substances for nourishment.

Those cells in growing plants that utilize carbon dioxide and water, along with solar energy, are autotrophic. They nourish themselves from CO_2 and in doing so are capable of storing up energy in the form of plant products such as carbohydrates, fats, and proteins. There are other autotrophic cells that use energy derived from chemical reactions as they feed on the simple materials in the environment. The source of chemical energy for these cells is frequently oxygen. In general, autotrophic organisms add to the storehouse of potential energy, and the energy-rich products of their activity become the source of energy for heterotrophic organisms. Thus, as heterotrophic organisms we must depend upon these products of autotrophic organisms, either directly through plant foods or indirectly through other heterotrophic organisms that feed on energy-rich autotrophic products. Ultimately, the energy we obtain from foods can be traced to the radiant energy (usually from the sun) stored up by autotrophic organisms.

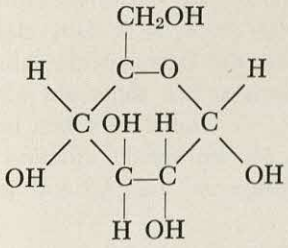
Living cells are the basic units of all organic processes, and, after examining the composition of several common types of organic substances encountered as cell components and cell products, we will return to some of the chemical processes that take place in living cells.

18-2 Carbohydrates

Carbohydrates, the most common foodstuff in the world, are composed of carbon, hydrogen, and oxygen. The name for this type of food comes down to us from the old days when it was first noted that the ratio of hydrogen to oxygen in these substances is 2 : 1, the same as that in water. These substances were thus called hydrates of carbon, or carbohydrates, although more recent studies of their composition show that this idea has nothing to do with their actual properties. The simplest carbohydrate is a sugar, and one of the commonest sugars is *glucose*, $C_6H_{12}O_6$. Chemical examination of glucose reveals that it contains five alcohol (hydroxyl) groups and an aldehyde group:



Glucose behaves both as an alcohol and as an aldehyde, though in acidic media the glucose molecule acts as a ring compound that has, as we can note, the same number of carbon, hydrogen, and oxygen atoms:

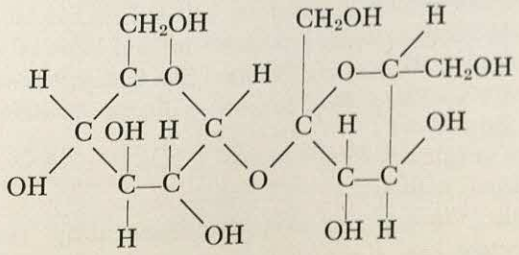


The glucose molecule, in the cyclic form, is the primary building block in starch and cellulose, two abundant components of plants.

One of the many isomers of glucose is *fructose*, which is the sweetest of the simple sugars. Fructose, or fruit sugar, may be combined with glucose to produce the common sugar, *sucrose:*

$$\underset{\text{glucose}}{C_6H_{12}O_6} + \underset{\text{fructose}}{C_6H_{12}O_6} \rightarrow \underset{\text{sucrose}}{C_{12}H_{22}O_{11}} + H_2O$$

Fructose has a ring structure similar to that of glucose, and the sucrose molecule contains these two rings attached to each other through an oxygen bridge:



Among other naturally occurring sugars similar to sucrose are maltose (malt sugar), gentiobiose (in gentian root), and lactose (milk sugar).

Starch is another common carbohydrate and is represented by the formula $(C_6H_{10}O_5)_n$. You will recall that when many small units are tied together in a giant molecule we have a polymer. Starch is a polymer composed of units of glucose. The simpler starch molecules contain about a hundred glucose units, whereas the more complex starches contain upward of several thousand glucose rings attached to each other through a series of oxygen bridges similar to the oxygen atom that holds the two simple sugar rings together in sucrose. By a simple experiment we can readily prove that starch is composed of glucose rings. When we boil starch in dilute hydrochloric acid, the big polymer is broken down first into dextrin (used for the mucilage on postage stamps), then into maltose, which is a sugar containing two units of the sugar rings, and finally into glucose. The over-all change, which is essentially a reaction of starch and water, is shown in the equation:

$$\underset{\text{starch}}{(C_6H_{10}O_5)_x} + xH_2O \xrightarrow{HCl} \underset{\text{glucose}}{xC_6H_{12}O_6}$$

We are all familiar with starch as a food that occurs in grains, potatoes, and many vegetables.

The most abundant plant product, *cellulose*, is a carbohydrate that is the essential component of the cell walls of plants—hence its name. About 60 per cent of wood is cellulose, over 90 per cent of cotton is cellulose, and the paper on which this book is printed is almost pure cellulose combined with clays and other fillers to make it smooth and opaque. Cellulose is again a polymer made up of glucose units, but, unlike starch, cellulose can be broken down into glucose only with a much more vigorous reaction. Cellulose is present in many foods, but, in contrast to starches and sugars, it is not digested—it merely provides bulk in our diet. In the next section we will see what happens to the simpler carbohydrates, such as sugar and starch, when they are digested.

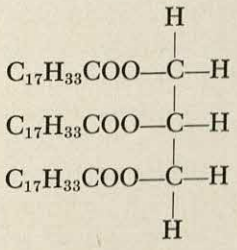
Before leaving cellulose, it may be interesting to note that it has, in spite of its apparent inertness, several chemical properties that lead to useful applications. The glucose rings in cellulose still have three hydroxyl or alcohol groups, and these may be converted into esters by treating cellulose with various acids. For instance, when cellulose reacts with nitric acid, an ester called *cellulose nitrate* is formed. Varying degrees of esterification of cellulose with nitric acid give us *guncotton* which is the base of smokeless gunpowders, and *pyroxylin*, which has fewer nitrate groups than guncotton and when dissolved in quick-drying solvents makes lacquers that are used on our automobiles and when mixed with camphor gives *celluloid*, an early, popular plastic. When cellulose reacts with acetic acid we get *cellulose acetate*, one

of the most common plastics for making inexpensive moldings, and thin fibers of it make one kind of *rayon* (celanese). The most common form of rayon, as well as cellophane, is derived from cellulose xanthate, another derivative of cellulose made by dissolving cellulose in sodium hydroxide solution and carbon disulfide (CS_2). This solution, which contains cellulose xanthate, is squirted through small holes into an acid bath where the original cellulose is regenerated. It is chemically the same as the original cellulose except that it is now in the form of a continuous filament. If the xanthate solution is extruded into the acid bath as a sheet, we get *cellophane.*

18-3 Lipids

Simple lipids are called *fats* and they constitute another type of foodstuff. Fats are esters composed of so-called fatty acids and glycerol. Typical fatty acids are palmitic acid, $C_{15}H_{31}$ COOH (palm oil), stearic acid, $C_{17}H_{35}COOH$ (beef fat), and oleic acid, $C_{17}H_{33}COOH$ (cottonseed oil). Although these esters are usually mixed—that is, a molecule of fat may contain two or three different acids attached to the glyceryl group—the predominance of one or two acids gives the fat a rather distinctive characteristic. For instance, many vegetable fats, such as olive oil, palm oil, and cottonseed oil are liquids, while animal fats, such as beef tallow and pork fat (lard), are solids. Let us consider two common fats, cottonseed oil and beef tallow.

Cottonseed oil contains chiefly the fat, glyceryl trioleate, and a molecule of this fat may be represented by the formula:



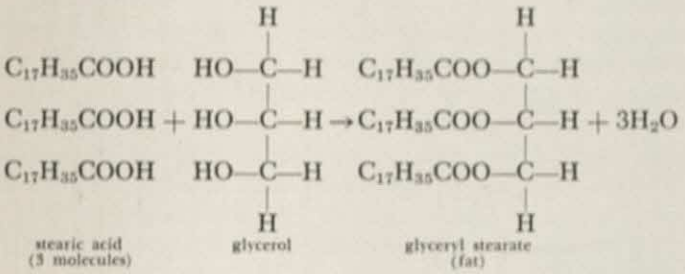
If we examine the oleate groups we find hydrocarbon chains similar to those in beef tallow, but with a notable exception: the oleate groups contain a double bond approximately midway in their carbon chain, and it is this unsaturation that is primarily the reason for this fat being a liquid. Americans usually prefer solid fats for food, and these liquid vegetable oils, such as cottonseed and corn oil, are converted into solid fats by a relatively simple process of hydrogenation:

$$\underset{\substack{\text{cottonseed oil}\\(\text{liquid})}}{(C_{17}H_{33}COO)_3{-}C_3H_5} + 3H_2 \rightarrow \underset{\substack{\text{margarine}\\(\text{solid})}}{(C_{17}H_{35}COO)_3C_3H_5}$$

This process, made possible by the presence of a nickel catalyst, produces a food product that is nutritionally and

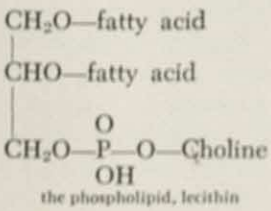
esthetically equivalent to animal fats, and, because of its lower cost, has become a popular source of fat in our diets. For instance, Crisco is hydrogenated cottonseed oil and Snowdrift is hydrogenated corn oil.

Beef tallow is mostly the ester glyceryl tristearate, and its structure can be verified by synthesizing it in the laboratory from stearic acid and glycerol:



The stearate groups contain long hydrocarbon chains of seventeen carbon atoms.

Complex lipids are important components of cells and, for example, may contain, in addition to fatty acids and glycerol, a phosphate group ($-HPO_4$) and a nitrogen-containing base such as choline or serine. The detailed composition of the lipids need not concern us here, except to gain an idea of the similarity between the simple lipids or fats, and the complex ones. A fat contains three fatty acids, but lipids contain only two fatty acids, and the third one may be replaced with the phosphate-base group:



the phospholipid, lecithin

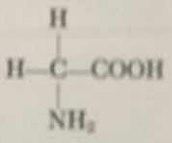
The cell membranes are apparently composed of phospholipids associated with proteins, and the uptake of nutrients by these membranes is crucial to the life processes of the cell.

18-4 Proteins and Amino Acids

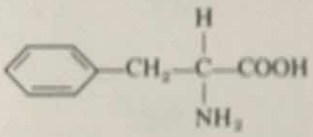
In Chapter 16 we became acquainted with two types of organic compounds: acids that contain the carboxyl ($-COOH$) group and amines that contain the $-NH_2$ group. We further saw that the acid and the amine groups may react with each other to form a linkage called the amide group, and you will recall that the amide group is responsible for the formation of the high polymer, nylon.

In a similar way, proteins are composed of various groups of carbon atoms held together by an amide linkage. If we take a natural protein apart by hydrolyzing it, we obtain the

acid and the amine groups that make up these amide linkages, and in the process we encounter a new kind of substance—an *amino acid*, which is an organic compound that contains both an amine group and an acid group. The simplest amino acid is *glycine:*



More complex amino acids may be derived from glycine by replacing one of the hydrogen atoms on the carbon with various organic groups. For example, an important amino acid in human proteins is *phenylalanine:*



In all, there are at least nine different amino acids that are essential in the human diet, and upward of thirty amino acids are known to occur in nature in plant and animal proteins. In addition to the amino acids, proteins contain small but important amounts of sulfur and phosphorus.

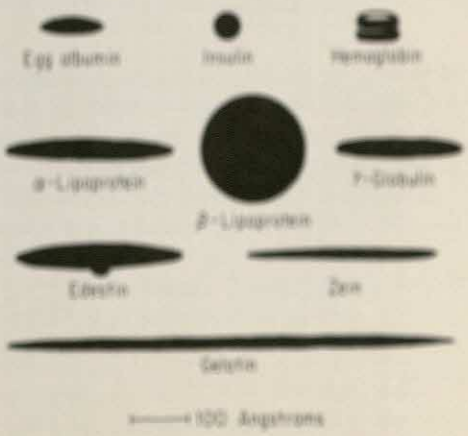
Protein molecules are of high molecular weight, the simplest ones being more than a hundred times the size of a fat or a sugar molecule. Not only are the protein molecules large, they have a complex physical structure that has aroused the curiosity of many scientists and which we will now briefly investigate.

A protein whose structure has only recently been determined is *insulin* (Fig. 18-1). This particular protein is a hormone that assists in the oxidation of sugars in the body and has a molecular weight of 6000. A single insulin molecule contains 51 amino acid groups attached to one another by amide linkages, and, as in a typical protein structure, the amine group of one amino acid is attached to the acid group of the next amino acid. By removing one amino acid after another, chemists have found out just what is inside an insulin molecule, but, like the schoolboy who dismantles his first alarm clock, they find it is quite another matter to put all the parts—in this case, the amino acids—together again in their proper places. Proteins are incredibly complex. Insulin is one of the few whose total structure is known at the present time. To get an idea of how many combinations of amino acids there can be, you must realize that the human body contains upward of 50,000 different proteins.

A single protein molecule may be spherical, rod-shaped, or an elongated ellipsoid (Fig. 18-2). If we were to unravel a protein molecule, we might obtain a linear chain 500 A long, and, at the same time, this protein, in its natural condition, may only be a sphere 25 A in diameter. Some pro-

Fig. 18-1. A model of the insulin molecule. The large white balls are sulfur atoms, the medium-sized white balls on the side groups are hydrogen atoms (many of these, for clarity, have been omitted), and the black balls are carbon atoms. The width of the molecule is about 20 Angstroms. (Model constructed by Paul Kitos)

Fig. 18-2. The shapes of several common protein molecules.



teins have been found to be made up of chains that are folded back and forth, or possibly coiled up much like a wire spring (Fig. 18-3). The protein that comprises hair and fingernails is such a coiled-up molecule. The mechanical strength of these coiled molecules is believed to be the result of hydrogen bonds between the —NH— groups in one spiral and the —CO— groups in a neighboring spiral. Just as the hydrogen bond accounts for the unusual properties of ordinary water, so here it explains some of the protein's special characteristics. For instance, the contraction and relaxation of the protein in muscle may be due to the breaking of certain hydrogen bonds and the reforming of others.

With the composition of these three types of biophysical substances—carbohydrates, fats, and proteins—let us now turn to some of the mechanisms by which organisms use up or produce these materials at the level of the living cell. First, let us examine the process of fermentation of carbohydrates.

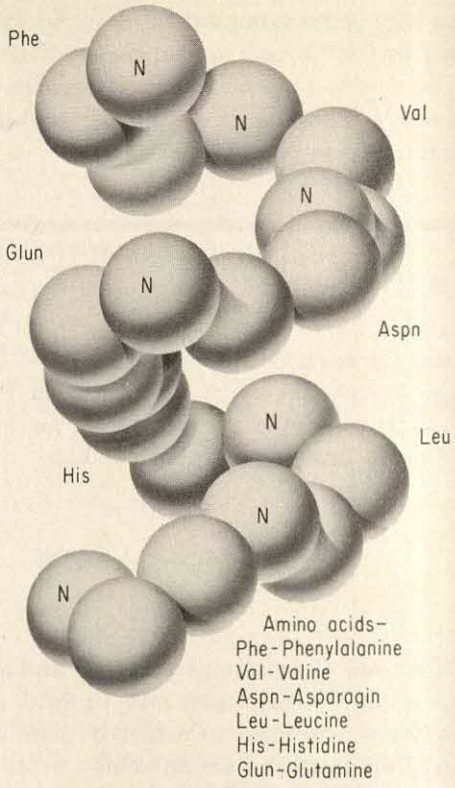


Fig. 18-3. Part of a coil of a protein molecule.

18-5 Fermentation

A biophysical process that has been used for centuries for converting complex carbohydrates to simpler substances is *fermentation*. This process is one of the oldest of chemical industries, very likely because the raw materials—molasses, fruit juices, and even wood, along with the requisite enzymes—are so readily available. For instance, when juice is pressed from grapes, the fruit sugars in the juice undergo fermentation because the enzymes required for this process are produced by yeast cells that adhere to the grape hull.

Ordinarily, in producing ethyl alcohol by fermentation, a special culture of yeast is added to a solution of molasses (a mixture of sugars). The yeast contains a mixture of enzymes called *zymase* which attacks the sugar and converts it into alcohol and carbon dioxide. For example, glucose would undergo the change:

$$\underset{\text{glucose}}{C_6H_{12}O_6} \rightarrow 2\underset{\text{ethyl alcohol}}{C_2H_5OH} + 2CO_2$$

However, for the large amounts of alcohol required by modern industry, fermentation methods of making ethyl alcohol are steadily giving way to cheaper ways that synthesize alcohol from petroleum by-products.

In order to understand the mechanism of a simple fermentation, which is the breakdown of glucose in the absence of air or oxygen, we must become acquainted with a fundamental reaction that is capable of carrying energy from one point to another in the living cell. This reaction involves a substance known as *adenosine triphosphate*, called ATP, for short:

$$\text{adenine} - \text{ribose} - P \sim P \sim P$$

Ribose is a simple sugar, the P's stand for phosphate groups and the adenine is a complex nitrogen base whose specific composition need not concern us here. The bonds indicated by ~ between the phosphate groups are high energy bonds because they absorb large amounts of energy upon formation. When they are ruptured they release about 8000 calories per mole. For example, adenosine diphosphate (ADP) is converted to adenosine triphosphate (ATP) by using up energy and holding it in a high energy P ~ P bond:

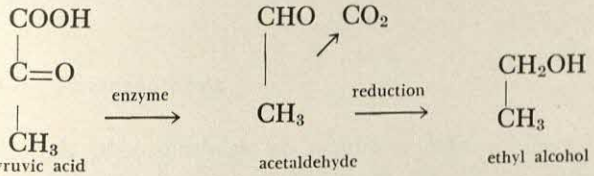
$$\text{ADP} + \text{inorganic P} + \text{energy} \rightarrow \text{ATP}$$

and conversely: ATP → ADP + P + energy (for use by the cell). ADP and ATP are called *coenzymes*, and, in conjunction with certain enzymes such as zymase, previously mentioned, these substances are critically important in the fermentation process.

In the fermentation of glucose to alcohol the ADP — ATP reaction is involved at least four times, sometimes adding energy, other times removing it. In the initial step, glucose is phosphated by ATP:

$$\text{glucose} + \text{ATP} \rightarrow \text{glucose-P} + \text{ADP}$$

An enzyme then converts the glucose-P to fructose-P which is further phosphated to fructose-P_2. This six-carbon substance is split into two three-carbon molecules which, after going through several additional intermediates, are attacked by an enzyme that splits off a carbon atom as CO_2:



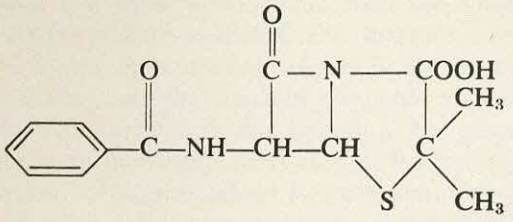
Other substances are involved in fermentation, but it is evident that ATP "sparks" the reaction. Because ATP is regenerated during the fermentation process, the reactions continue as long as there is glucose and the enzymes are capable of performing their functions. Eventually, the concentration of alcohol in the "brew" reaches a point where the enzymes can no longer function and fermentation ceases. Fermentation can be performed in cell-free media, but it is more rapid in the presence of living yeast cells.

If we wish to break down large starch molecules by fermentation, another enzyme particularly suited for this purpose, called *diastase*, is needed. Diastase occurs in the sprouts of barley seeds (the malting process) and accounts for the large quantities of barley that are used in the brewing industry. Each carhohydrate is best fermented by a special enzyme, and, as a result, there are a host of different kinds of enzymes employed in the fermentation process.

In addition to alcohol fermentation, enzymes bring about the synthesis of other substances. They account for the formation of vinegar from apple juice and lactic acid in sour

milk, and those produced by mold growth give us many fine cheese flavors, such as Roquefort and blue cheeses. In industry, enzymes are essential in the manufacture of butyl alcohol and acetone, two widely used organic solvents for paints and lacquers.

One spectacular example of mold synthesis of particular modern interest in *penicillin.* Although penicillium molds had been known for many years—Pasteur in France in 1858 found that these molds would differentiate between the two isomers of tartaric acid—it remained for Sir Alexander Fleming of England to discover in 1929 that penicillium molds contain an antibacterial substance. In his laboratory one day Fleming was studying culture plates of common bacteria of the coccaceae family—of which pneumonia and streptococcus are members—and by chance a stray mold moved in on one of the cultures and stopped its growth. Fortunately, Fleming was prepared to make the most of the accident; he isolated the mold, found it to be penicillium, and, as a result, introduced the first practical antibiotic. When the substances in the mold were later separated, it was found that the chemical substance responsible for killing the bacteria is a single, complex organic substance—penicillin. Although it yielded the secret of its structure only after several years' research by many scientists, we can see in its formula several familiar groups—the benzine ring, two smaller rings that contain nitrogen and sulfur atoms, and an acid group:



penicillin

Fabulously expensive during World War II, the cost of penicillin has today fallen to only a few cents a dose because the methods of growing the mold that makes this enzyme have been greatly improved. However, a cheap method of synthesis has still not been found, and, unlike our ability to produce ethyl alcohol by more convenient means than fermentation, we will likely have to depend on improved strains of penicillium mold to produce penicillin for many years. So this is indeed an exciting frontier in biophysical science.

18-6 Respiration

The fermentation of glucose occurs in the absence of oxygen and is called an *anaerobic process.* Respiration, on the other hand, uses oxygen, and the over-all reaction for the complete respiration of glucose may be represented:

$$\underset{\text{glucose}}{C_6H_{12}O_6} + 6O_2 \rightarrow 6CO_2 + 6H_2O + \text{energy}$$

This is a stepwise process in living cells, and the energy is released in small amounts along the way. If the energy were released in one large burst, so much heat would be released that the temperature would rise to the point where it would exceed that tolerated by living organisms. Fortunately, respiration is a comfortably slow process.

When we eat glucose or other digestible carbohydrates such as other sugars and starch, much of the glucose obtained from them is not burned up immediately, but large amounts of it are stored up as glycogen (animal starch). In an adult, over 100 grams of glycogen may be stored in the liver and 250 grams may be stored in muscle tissue. In plants, glucose is stored up by chemically converting the glucose formed in photosynthesis to carbohydrates such as cellulose or starch.

In terms of over-all efficiency, the cell is about 55 per cent efficient in converting the energy in glucose to a useful form of energy in the form of ATP which the cell may use for its work. Since ordinary machines such as gasoline engines rarely operate with an efficiency greater than 25 or 30 per cent, the cell is uniquely equipped to supply us with a large portion of the energy in our foods. It is of interest to note that cells metabolize other foodstuffs such as fats and proteins, but the mechanisms for these processes are beyond our concern in this book.

18-7 Photosynthesis

Through *photosynthesis* an estimated billion tons of organic substances burned up by living systems are regenerated daily. The autotrophic cells in plants put in a long day converting carbon dioxide and water into carbohydrates, fats, proteins, and other organic products in plants. The energy for this huge task is provided by the sun, but before considering some of the details, let us explore some factors that enable us to better understand and appreciate this most unique and interesting of physical science topics. In Chapter 12 we became acquainted with the quantum concept of energy and saw that light emitted by an excited atom reveals the energy states within the atom. Further, we saw that light itself is a stream of tiny bundles of energy, called photons, and that each photon possesses energy that is proportional to the frequency of the light. The energy of each photon, or quantum of energy, is given by the equation (Equation 12.1):

$$E = hn$$

The energy of a photon may be sufficient to initiate certain chemical reactions. For example, a mixture of hydrogen and chlorine gas may be kept in a glass container for some time

in the dark, but if this mixture is exposed to sunlight, a violent explosion occurs. We may describe this reaction by the equation:

$$H_2 + Cl_2 + \text{photons} \rightarrow 2HCl + \text{energy}$$

The photons supply the necessary activation energy, and once the reaction is started, the energy produced by the exothermal change is sufficient to sustain the reaction until one or the other of the reactants is used up.

Two important laws that apply to the effect of light upon chemical processes are relevant here. One of these states that *only the light that is absorbed can be effective in initiating a reaction.* Thus, at least one of the substances in the reaction mixture must be colored so that it will absorb light. In the hydrogen-chlorine reaction, the greenish-yellow color of the chlorine is responsible for absorbing light, and in plants the green chlorophyll absorbs light. The second law says that *one photon is sufficient to start one molecule reacting.*

If we apply this second law to an important reaction that occurs in nature, we will find that when light of very short wave length (ultraviolet light) passes through oxygen, some of it is absorbed—that is, oxygen molecules, although transparent to visible light, are capable of absorbing light between 1600 and 1800 A wave length—and for each photon of light energy absorbed, an oxygen molecule is split in two:

$$O_2 + \text{photon} \rightarrow 2O$$

This is spontaneously followed by the reaction:

$$\underset{\text{oxygen atom}}{O} + \underset{\text{oxygen molecule}}{O_2} \rightarrow \underset{\text{ozone}}{O_3}$$

The ozone may eventually react with one of the oxygen atoms:

$$O_3 + O \rightarrow 2O_2$$

Hence, in the atmosphere about 15 miles above the earth's surface, there is an ozone layer in which these reactions absorb practically all the ultraviolet light that comes to the earth from the sun. If it were not for this protective layer, life processes as we know them on the earth would not exist, because ultraviolet light is severely destructive to the many organic substances that support life—including our own bodies. We can also see the effects of light energy in such processes as the fading of dyes in sunlight, sunburn of the skin, the darkening of silver salts in photographic film, and the making of blueprints.

The most important chemical process that is initiated by light energy is *photosynthesis.* We know that green plants require the presence of sunlight in order to grow, and we also know that green plants are responsible for keeping the atmosphere fresh, by regenerating oxygen from the carbon dioxide that is continuously being added to the atmosphere. The over-all reaction in photosynthesis is the conversion of

carbon dioxide and water into carbohydrates and oxygen, in the presence of chlorophyll. This may be shown briefly by the equation:

$$6CO_2 + 6H_2O + \text{energy} \rightarrow \underset{\text{glucose}}{C_6H_{12}O_6} + 6O_2$$

Glucose, as we know, is a representative carbohydrate to which many of the more complex carbohydrates, such as starch and cellulose, are related. It is interesting to note that the energy we derive from these foods is essentially the energy of sunlight stored up by means of the endothermic process of photosynthesis. It is believed that one photon of energy is sufficient to start one molecule of carbon dioxide on its way to forming glucose.

The photosynthesis process is by no means as simple as implied in the equation. Nor are all the details of the photosynthesis process understood and agreed upon by investigators in this area of study. One of the favorite procedures for studying photosynthesis makes use of one of the simplest plants—green algae—by exposing it to light for short intervals. The reactions are stopped and the intermediate steps are examined. Much of the mystery of photosynthesis was resolved by Melvin Calvin and his associates at the University of California, using this procedure.

In addition to the familiar ATP — ADP pair, photosynthesis involves another pair of substances, namely *diphosphopyridine nucleotide,* called DPN^+, and its partner DPNH. In the initial step of photosynthesis (see Fig. 18-4), chloro-

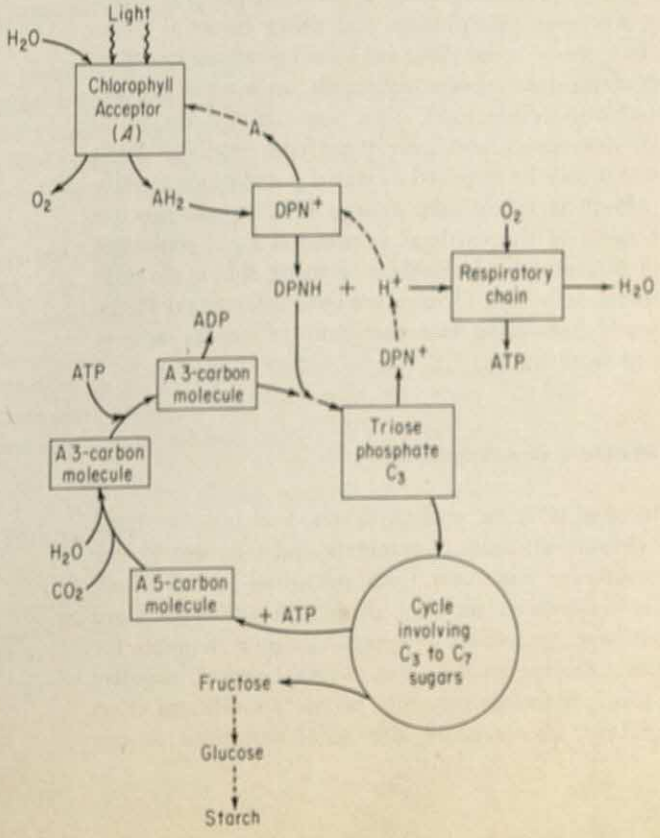


Fig. 18-4. Some of the steps in photosynthesis.

phyll makes possible the splitting of water into hydrogen and oxygen. The oxygen is evolved as gas and the hydrogen is carried by an acceptor, A (which may be chlorophyll itself), to DPN^+:

$$DPN^+ + AH_2 \rightarrow A + DPNH + H^+$$

The DPNH then enters a cycle as shown in Fig. 18-4. It is interesting to note the several places where the ATP — ADP pair injects energy into the process and that the high-energy ATP required for these steps is derived from a respiratory side chain. Light energy is required only for the initial steps in photosynthesis, and once the DPNH is formed, the succeeding steps proceed in the dark. The traditional observation that one can hear the corn grow in a field at night may be attributed to those parts of the photosynthesis reactions that do not require light energy.

The chief product of photosynthesis is the class of substances called *carbohydrates*. Many other plant products, such as proteins, fats, phospholipids, and vitamins, are synthesized in reactions related to photosynthesis. The mechanisms for many of these have been studied, but a consideration of them is much too complex to include here.

As we learn more about the mechanism of photosynthesis we will have a better chance to harness this process and perhaps to improve it. Plants apparently house the most efficient means of converting solar energy into other forms of energy, and since algae are one of the most efficient plants in this respect, we might improve considerably on conventional methods of producing foods by growing algae in tanks containing nitrogen, phosphorus, and other essential plant nutrients. In contrast to conventional farming, where less than 1 per cent of the solar energy falling on an acre of ground is converted into edible food, algae can capture up to 20 per cent of this energy and store it in food products. Such a development may be required as world populations rapidly increase. About as much solar energy falls in one day on 75 square miles of the earth as is released by a hydrogen bomb, and the most likely method of using this is through photosynthesis. In a later chapter we shall refer again to the possibilities of harnessing this vast store of energy and of putting it to work for us.

18-8 Minerals and Vitamins

In addition to proteins, carbohydrates, and fats, we need small but definite amounts of minerals and vitamins in our diet. Our needs for these vary from individual to individual, and certain difficulties, such as alcoholism and premature aging have been traced to a person's greater demand for some of these substances than the so-called normal requirements. In much the same way that people are tall and short or fat and thin, so there are also wide variations in our

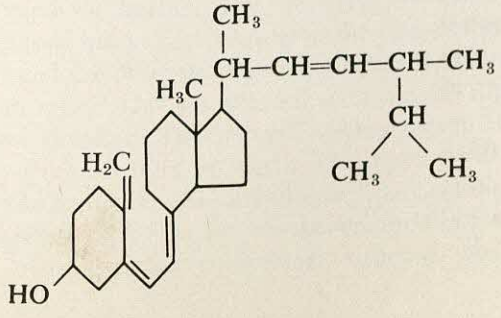
needs for vitamins and minerals. The minerals required in our diets include compounds of sodium, potassium, calcium, magnesium, iron, and copper and also chloride, iodide, and phosphate compounds. From sodium chloride, for instance, comes the hydrochloric acid for the gastric juice in the stomach, and iron is needed primarily to form the red hemoglobin in the blood. Iodine is used to synthesize thyroxine, an amino acid essential to thyroglobulin, which is a protein that controls the rate of growth and metabolism in our bodies. Calcium and phosphorus are needed to make teeth and bones. Even traces of copper, cobalt, and molybdenum are necessary, and for this reason we should not limit our diet to the most highly purified foods because we may not get these important traces.

Of more complex nature are the *vitamins*. It has been known for a hundred years that certain foods contain substances that prevent such diseases as scurvy, rickets, pellagra, and beriberi. In recent years, small quantities of these special substances, or vitamins, have been detected in our foods, and, although we usually eat enough of them, occasionally it is advisable to supplement our diet with vitamin pills or capsules.

Our bodies require at least thirteen known vitamins, but because of their complex chemical nature, we will consider only a few representative ones. Vitamin A, an alcohol whose formula is $C_{20}H_{29}OH$, is an oily, yellow substance that is derived from the red and yellow pigmented material in tomatoes, carrots, green vegetables, eggs, and milk. It prevents skin disorders and infection, and is apparently responsible for supplying our eyes with the substances required for night vision.

There are at least four essential B vitamins, and they contain nitrogen or amine groups that suggested to early workers that these are "vital amines," hence the origin of the word "vitamin." The most recently discovered vitamin in this group, vitamin B_{12}, assists in the production of red corpuscles and in the control of disease. We need only 1×10^{-6} g of this vitamin per day, but, with a molecular weight of about 1400, including an atom of cobalt, it possesses by far the most complex structure of all the vitamins thus far discovered.

The vitamin that is necessary to preserve healthy bones and teeth is vitamin D, and to appreciate its complexity let us inspect the chemical structure of one form of it, vitamin D_2:



The structure of this vitamin, in view of our limited introduction to organic compounds in Chapter 16, may be a bit confusing, but if we examine this formula, we will find that the vitamin consists of an alcohol group, several methyl groups, and typical carbon-carbon bonding. It is beyond our interest at this point to retrace the steps that were taken to find this formula, but the study of these complex substances has captured the interest of many scientists, and, since their methods and techniques have been based upon sound concepts, we may have confidence in their results. In fact, so dependable have these results been that we often prefer the substance produced in the laboratory to that produced in nature. This is true in the case of drugs and medicinals, which we will now consider, as well as of vitamins.

18-9 Chemotherapy

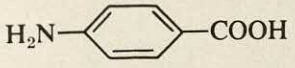
Along with the normal processes of metabolism of ordinary foods, it is appropriate to consider here some of the biophysical aspects of abnormal processes such as disease. The practice of treating diseases by means of special substances has been pursued for centuries, and this practice is called *chemotherapy*—the treatment of disease with chemicals.

Even in primitive civilizations there were favorite remedies for the prevalent diseases. Some of these were undoubtedly effective, others not, but through trial and error a store of knowledge regarding the therapeutic effects of many of these substances was gradually accumulated. Although the extracts of certain roots, barks, and herbs have long been favorite sources of medicants, only in recent years, with the application of scientific analysis, have we learned of their effective components. For example, the bark of the cinchona tree in Central America was found to provide an extract that is a specific remedy for malaria. The extract is *quinine*, a chemical substance having the formula $C_{20}H_{24}N_2O_2$. Through careful studies of the structure of quinine, scientists have succeeded in synthesizing many similar substances that are even more effective for the treatment of malaria and similar diseases.

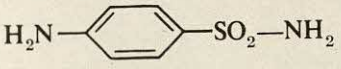
A good example of the application of modern research techniques to chemotherapeutic substances is the research that led to the discovery of "sulfa" drugs. Early in the twentieth century scientists noted that certain dyes, such as crystal violet and methyl violet, were effective in the control of certain bacteria. Later, through studies of the structure of these dye molecules, they found that certain groups of atoms in these dyes were responsible for the bacterial action. With this in mind, new compounds containing these groups were prepared in the hope of finding a more potent bactericide, and, oddly enough, the most effective compound turned out to be a much simpler one than the original dye.

In contrast to methyl violet, which has the complex formula $C_{25}H_{30}N_3Cl$, *sulfanilamide,* with a relatively simple formula of $C_6H_8O_2N_2S$, became the first really successful sulfa drug. It was immediately put to wide use because of its rapid action upon pneumococci, streptococci, and similar bacteria, and during World War II it greatly lowered the number of fatalities from wounds by preventing infection by bacteria.

The mechanism by which sulfa drugs kill bacteria is an exciting story in the progress of science. The "sulfas" do not wage a frontal attack on bacteria; they trick the bacteria into a trap. Bacteria regularly feed upon a substance called *para-aminobenzoic acid,* which is necessary for their growth, and, since the sulfa drugs have a structure quite similar to that of para-aminobenzoic acid:



para-aminobenzoic acid



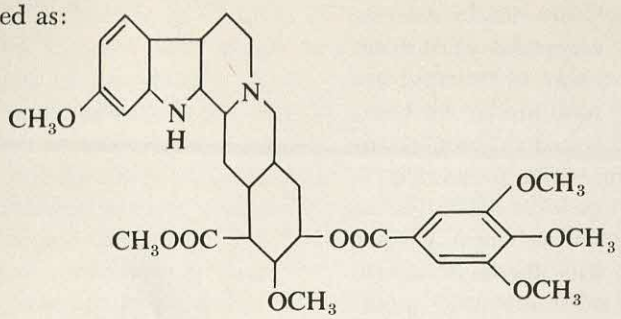
sulfanilamide

the bacteria will absorb either of these substances. The sulfa compound, however, cannot be utilized by the bacteria and therefore it stops their growth, prevents them from reproducing, and allows the normal body functions of the host to take over and defend the body against the bacteria.

Even more effective than the sulfas are the so-called broad spectrum antibiotics, so named because they are effective in the treatment of a broad group of bacteria or infectious organisms. One of these is penicillin, which we have already examined. The pre-eminent success of penicillin in chemotherapy has fostered a widespread search in recent years for other naturally occurring substances of this type, and, as a result, we now have *streptomycin, aureomycin,* and *chloromycetin* for treating bacterial and virus infection. Another potent antibiotic is *terramycin,* derived from the soil, and although most of these have been synthesized it is usually more economical to obtain them through a natural process.

In their persistent search for substances that may benefit man, scientists are continuing to add new ones to the list of chemotherapeutic agents. The discovery of two complex substances, one occurring naturally in the juices of the snakeroot plant and the other produced synthetically, give promise of relief to many who are afflicted with mental illness and high blood pressure. The naturally occurring substance is *reserpine,* crude extracts of which have been used for centuries by the medicine men of India, but only recently has it been identified as the active component in the root juices used for treating insanity, insomnia, epilepsy, and related

disorders. Once separated from its natural source, reserpine has been identified as:



The synthetic compound is *chloropromazine,* and it and reserpine are called *tranquilizers.* Together they have ushered in a revolution in the treatment of mental illness and hold considerable promise for treatment of insomnia, alcoholism, and obesity.

18-10 Fossilization of Carbohydrates

In the geological realm, a biophysical process that is primarily concerned with carbohydrates is the conversion of carbohydrates into fossils. When plants, or parts of plants such as leaves, die they undergo one of two changes. If exposed to air and acted upon by enzymes, they may be fermented and eventually converted into carbon dioxide and water. On the other hand, if they accumulate in swampy regions they may be transformed into various kinds of coals by a process known as *fossilization.* Although coals have been formed in more recent times, it was during the late Paleozoic era, some 200 million years ago, that conditions were most favorable to fossilization of plants, and most coals of the Appalachian and Midwest regions were formed at that time.

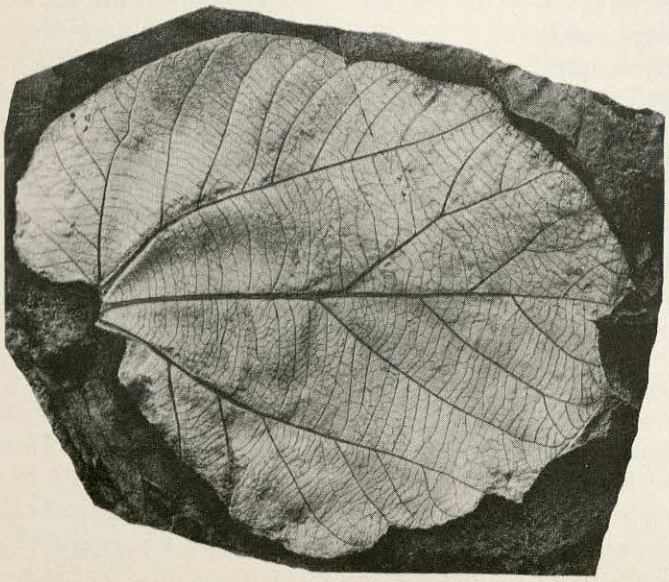


Fig. 18-5. The imprint of a leaf as a fossil in coal found near Trinidad, Colorado. (U. S. Geological Survey)

A fossil is an imprint, a trace, or the actual remains of a living organism that has been deposited during the geological past. The footprint of an animal now preserved in rock, the skeleton of an animal, an insect captured in a resinous deposit, petrified wood—all these are fossils. That coal is a fossilized plant material is indicated by two observations. First, coal often contains the imprints of leaves and stems of plants that grew long ago in broad, swampy regions (Fig. 18-5). Second, the composition of coals shows a gradual change as they are stored for longer and longer periods of time. Referring to Table 18-1, which lists materials in the order of their age, we note that the percentage of carbon increases while the percentage of hydrogen and oxygen decreases. The composition of the original wood, and of each of the substances in the chain from wood to graphite, is altered in the following manner: the volatile elements such as hydrogen and oxygen, which are probably in the form of water, are lost, while carbon, which is nonvolatile, remains behind. Obviously, long periods of time and favorable conditions of temperature and pressure are necessary for these changes.

TABLE 18-1 COMPOSITION OF SUBSTANCES INVOLVED IN FOSSILIZATION OF CARBOHYDRATES

	Carbon	*Hydrogen*	*Oxygen*
Wood	50%	6%	44%
Peat	59%	6%	35%
Brown coal	69%	6%	25%
Bituminous coal	82%	5%	13%
Anthracite	95%	2.5%	2.5%
Graphite	almost 100%		

Most of the substances in Table 18-1 may be viewed as sedimentary rocks that are laid down and gradually covered by later sediments. After being subjected to the action of increased pressures and temperatures, ordinary soft coals, such as bituminous coal, are converted into hard, or anthracite, coal. Since this is essentially a metamorphic process, anthracite is a metamorphic rock.

We should not leave the topic of fossilization without bearing in mind that plants contain other organic materials in addition to carbohydrates. Coal, for instance, contains small amounts of nitrogen and sulfur, and these elements are found in many of the nearly 200 organic compounds that can be recovered from coal. The recovery of these compounds from coal is a big industry, and because of the large deposits of coal available we should have a dependable source of these useful materials for many years.

18-11 Summary

The unit of living processes in plants and animals is the

cell. Cells capable of nourishing themselves directly from the environment are autotrophic and those that must depend upon organic substances for food are heterotrophic.

Carbohydrates include sugars, starch, and cellulose, with glucose being somewhat the common denominator of them all. Carbohydrates are the most common substances in the biophysical world. They are stored as cellulose in plants and as glycogen in animals, and they supply energy through respiration or fermentation processes.

Lipids may be simple (fats) or complex (such as phospholipids) and they serve as a storehouse of energy and as important components of cells. Fats are glyceryl esters of fatty acids, and the complex lipids contain in place of a fatty acid a phosphate group and an organic base.

Proteins are composed of amino acids attached to each other through amide links. Proteins in our diet are needed to repair and maintain the muscle, tissue, skin, hair, and nails.

Fermentation is the enzyme-catalyzed decomposition of organic substances, frequently occurring in the absence of oxygen. When oxygen is present the enzymes support a process called respiration which burns the substances to CO_2 and H_2O.

Photosynthesis is the process by which solar energy is stored up in energy rich substances such as the various plant products. The main materials for this process are carbon dioxide, water, phosphorus, nitrogen, and less common substances, and the photosynthetic reaction is "sparked" by several complex substances such as chlorophyll, DPN^+ and ATP.

Vitamins and minerals, though required in small amounts in the diet, are essential to regulate normal cell functions. The treatment of disease with chemical substances is called chemotherapy. Both natural and synthetic substances have been found to be effective for certain diseases.

Fossilization of plant products is a geological process that results in the formation of peat, lignite, and coal, which contain progressively higher amounts of carbon and lower amounts of hydrogen and oxygen.

EXERCISES

A. KEY TERMS TO REMEMBER

ADP	enzyme	penicillin
amino acid	fat	photosynthesis
ATP	fermentation	protein
autotrophic	fossil	respiration
carbohydrate	glucose	starch
cellulose	heterotrophic	sulfa drug
DPN^+, DPNH	lipid	vitamin

B. QUESTIONS ABOUT BIOPHYSICAL SUBSTANCES

1. What are the differences between autotrophic and hetero-

trophic organisms in the sources of nourishment and their metabolic products?

2. List the primary differences between glucose, sucrose, starch, and cellulose.

3. How are guncotton, rayon, cellophane, and pyroxylin derived from cellulose?

4. What is the chemical difference between a liquid fat such as cottonseed oil and a solid fat such as stearin?

5. Explain in general how some thirty amino acids may form more than 50,000 different proteins in the human body.

6. List several natural and synthetic processes that are essentially fermentation processes.

7. When we burn wood in the fireplace, we obtain a certain amount of heat. Could we expect to obtain about the same total amount of heat for our bodies by the consumption of an equivalent amount of glucose or starch? What is the fundamental difference between these two processes?

8. State two laws that have been derived from the effect of light upon chemical reactions.

9. List several changes that might occur on the earth if the ozone layer in the upper atmosphere did not absorb the major portion of the ultraviolet light coming from the sun.

10. Discuss the role of the ATP — ADP pair in fermentation, respiration, and photosynthesis.

11. What is the specific role of chlorophyll in photosynthesis? What is meant by the "dark reactions" in photosynthesis?

12. If a piece of filter paper (cellulose) is placed in moist soil for a few days, then removed and examined, it will be found to contain some glucose. What is the source of the glucose?

13. A thermos jug is filled with wheat or barley and the seeds are moistened with water. A thermometer is inserted in the jug and observed frequently over a period of several hours. The temperature is noted to rise. Explain the source of this heat energy.

14. Design a simple experiment by which you could prove that light striking a green, growing plant leaf causes the evolution of oxygen. Hint: Oxygen is only slightly soluble in water.

C. PROBLEMS ABOUT BIOPHYSICAL SUBSTANCES

1. Calculate the molecular weight of glucose, glyceryl tristearate, and penicillin. Ans. 180.

2. About how many glucose units would be in a starch whose molecular weight is approximately 15,900?

Nineteen

Electrochemical Energy

Several times in earlier chapters we have encountered the concept that matter and electricity are closely related to each other. Electricity is a flow of electrons; matter is made up of atoms whose outer structures consist of electrons. It has been pointed out frequently that when substances react chemically with one another there is a rearrangement of the outer electrons of the atoms that are in the reaction. Hence, chemical reactions also involve a movement of electrons; indeed, we might say that all chemical reactions may be viewed as electrical phenomena. Even the prosaic combustion of coal can be considered as essentially an electrical process—not, however, to be confused with the process by which coal is burned to supply the heat used to drive steam turbines (electrogenerators). In fact, several partially successful attempts have been made to generate electricity directly from coal as it undergoes a chemical reaction. Let us examine more closely the electrical energy that may be obtained from chemical processes.

Whenever we place a piece of metal in water, the metal tends to go into solution. Some metals, such as sodium and calcium, dissolve so readily that they react quite vigorously with the water. Other metals, such as iron, lead, and copper, show less tendency to dissolve. Although when we place these last three metals in water we might conclude that there is no reaction whatsoever, a small amount of these metals do dissolve, even though the process is not visible to the naked eye. The sensitive electrical instruments available to the scientist can detect even the small amount—and it is indeed slight—of silver and gold that dissolves in water.

Before going further, it may be well to refresh in our minds the meaning of the electrical terms: volt, coulomb, ampere,

and ohm. The most important of these in our present discussion is the volt.

When one joule of work is needed to move one coulomb of electrical charge between two points, a difference of potential of one volt exists between these points. That is, when one coulomb (6.23×10^{18} electrons) of charge drops in potential by one volt, it can do one joule of work for us. The volt represents a potential energy per unit of charge, that is, the energy difference between the two points, and this concept is in keeping with the idea of potential mechanical energy, for that energy is proportional to the mass and to the height (difference in elevation).

19-1 Electrochemical Energy

In the discussion that follows, electrochemical energy is expressed with respect to two points—one before a reaction has occurred and the other after it has occurred. For example, when metallic sodium is placed in water, the sodium atoms become sodium ions according to the equation:

$$\underset{\text{atom}}{Na} \rightarrow \underset{\text{ion}}{Na^+} + \underset{\text{electron}}{e^-}$$

We may say that the sodium atom has a certain amount of electrochemical energy and that the sodium ion together with the electron have a different amount of electrochemical energy. The difference between these two amounts of energy is a potential difference, or a certain voltage.

The most precise method of measuring the voltage in the ionization of a sodium atom is to place gaseous sodium atoms in a special vacuum tube in which the sodium atoms are made to dissociate by bombarding them with electrons. The bombarding electrons are energized from an outside source of voltage, the amount of which gives us a measure of the energy required to pull an electron away from a sodium atom. It requires 5.14 electron-volts to remove an electron completely from a sodium atom, and this method of measuring the ionization potential of various atoms provides experimental proof of the existence of energy levels in atomic structure (Chapter 12). However, these experiments are limited to gaseous atoms. Since we are presently more concerned with reaction in solutions, let us examine the methods of measuring the tendency for metals to go into solution.

Suppose we place a piece of zinc in a solution of zinc sulfate and a piece of copper in a solution of copper sulfate (Fig. 19-1). Between the two solutions is a porous partition which prevents the solutions from mixing with each other, but which permits ions to migrate from one solution to the other. When the zinc and copper are attached to a voltmeter, we find a potential of 1.10 volts between these two metals. Where does this potential arise? An examination of this system will enable us to find the mechanism by which it operates.

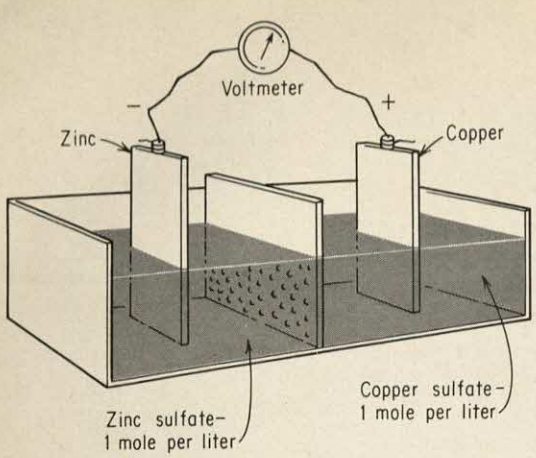


Fig. 19-1. A zinc-copper electrolytic cell.

The zinc is negative and the copper is positive, that is, there is an excess of electrons on the zinc, a lack of electrons on the copper. Since electrons on the zinc have 1.10 volts greater potential energy than do electrons on the copper, electrons will flow through the outside circuit from the zinc to the copper. If we wish, we may make use of this energy by placing a suitable motor or a lamp between the zinc and copper. If we let the electrons flow from the zinc to the copper for an extended period of time, we will note three results: first, the piece of zinc will become smaller because some of it will dissolve; second, the piece of copper will become larger; and third, the copper sulfate solution will lose some of its intense blue color. This last result tells us that the copper is leaving the solution and plating out on the piece of copper. For each copper atom that plates out, a zinc atom goes into solution. This is possible because zinc loses two electrons to form a zinc ion (Zn^{++}) while the copper ion (Cu^{++}) gains two electrons to form a copper atom. The over-all process is described in the diagram in Fig. 19-2. Since the loss of electrons is an oxidation process and the gain of electrons is reduction, the zinc becomes oxidized and the copper becomes reduced. The basic reason why this is a spontaneous process lies in the fact that electrons possessed by zinc have more energy than electrons possessed by copper, and this fact is the driving force in the transfer of electrons from the zinc to the copper.

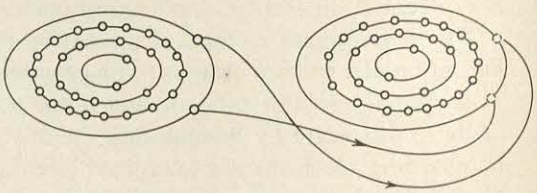


Fig. 19-2. In the zinc-copper cell, the zinc becomes oxidized and the copper ion becomes reduced.

There is no way that we can experimentally measure the tendency of a single metal to go into solution. We must always compare the solution tendency of one metal to that of another. In the above example, we examined the solution tendency of zinc compared to that of copper and found further evidence that an oxidation process is accompanied by a reduction process. Indeed, one of these processes can never occur without the other. As a standard reference, the oxidation of hydrogen has been selected as the basis for comparing the measurements of the electrochemical potentials of all the elements. When we compare the tendency of hydrogen to dissolve to the tendency of zinc to dissolve, we find that zinc's tendency is the greater. That is, zinc loses

two electrons with 0.76 volt greater energy than does hydrogen. If we say, arbitrarily, that the hydrogen potential is zero, then the zinc potential compared to hydrogen is 0.76 volt (Fig. 19-3). In all the comparative measurements with hydrogen, the concentration of the solution, the temperature, and the pressure of the hydrogen should be at the standard conditions shown in the figure.

If a metal shows less tendency to go into solution than does hydrogen, then its electrochemical potential is expressed as a negative voltage. A list of the electrochemical potentials of the elements is useful in describing the solution tendency of the elements as well as to define their chemical reactivity in general. From such a list (Table 19-1), we can find that the voltage of a standard magnesium-copper cell would be 2.34— (—0.34) volts, or 2.68 volts. This voltage would provide ample energy to operate a flashlight lamp, as is demonstrated in Fig. 19-4. Since the conditions of the

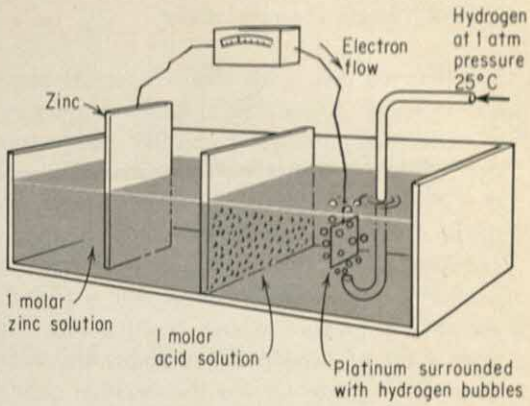


Fig. 19-3. The measurement of the standard electrochemical potential of zinc by comparing it to the standard hydrogen cell.

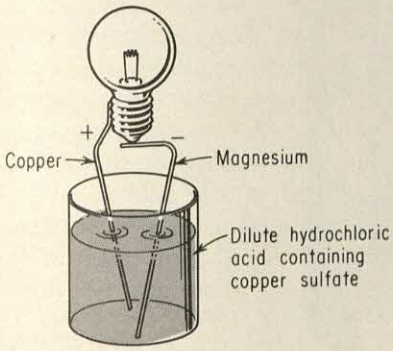


Fig. 19-4. A magnesium-copper cell.

cell shown in the figure are not standard, the voltage will not be that predicted from Table 19-1, but it will be sufficient to light up the small bulb for a short time until the magnesium is used up.

TABLE 19-1 STANDARD ELECTROCHEMICAL POTENTIALS

Measured at room temperature in solutions containing 1 mole of the ion per liter of solution

Reaction		*Voltage*
$K \rightarrow K^+$	$+1e^-$	2.92
$Na \rightarrow Na^+$	$+1e^-$	2.71
$Mg \rightarrow Mg^{++}$	$+2e^-$	2.34
$Al \rightarrow Al^{+++}$	$+3e^-$	1.67
$Zn \rightarrow Zn^{++}$	$+2e^-$	0.76
$Fe \rightarrow Fe^{++}$	$+2e^-$	0.44
$Pb \rightarrow Pb^{++}$	$+2e^-$	0.13
$H_2 \rightarrow 2H^+$	$+2e^-$	0.00
$Cu \rightarrow Cu^{++}$	$+2e^-$	−0.34
$Hg \rightarrow Hg^{++}$	$+2e^-$	−0.85
$Pt \rightarrow Pt^{++}$	$+2e^-$	−1.2
$Au \rightarrow Au^+$	$+1e^-$	−1.68

19-2 Electrochemical Cells

When we deal with electrochemical processes, there are a few terms that we should know in order to understand the discussion that follows. The two conductors that form the positive and negative poles are called *electrodes*. The positive electrode is usually called the *anode*, the negative electrode the *cathode*. In an electrochemical cell, oxidation occurs at the cathode, reduction at the anode. The liquid portion of an electrochemical cell is called the *electrolyte*, an ionic substance whose positive ions (*cations*) migrate toward the negative pole (cathode) and whose negative ions (*anions*) migrate toward the positive pole (anode). Sometimes the reaction that occurs at one of the electrodes is called a half-cell reaction because two of these reactions—one oxidation and the other reduction—are needed to make a complete electrochemical cell. Table 19-1 is essentially a list of half-cell reactions.

Almost all chemical reactions that involve oxidation and reduction can be adapted to create an electrochemical potential in a process in which the oxidation reaction produces the negative pole and the reduction reaction the positive pole. However, most of these chemical processes would be too expensive or too inconvenient for practical use. Only a relatively small number of these reactions have been successfully adapted to the production of a convenient, portable source of electrical energy. Let us examine two common examples.

The so-called *dry cell* is an electrochemical cell that produces approximately 1.5 volts of electrical energy (Fig. 19-5). We are all familiar with the ordinary flashlight cell that provides us with a convenient source of electrical energy. The active components of the dry cell include the zinc can that serves as the negative pole, a paste containing ammonium chloride and manganese dioxide, and a carbon rod that serves as the positive pole at the center of the cell. The chemical reactions that take place when the cell is being used are not fully understood, but it is quite certain that the zinc goes into solution, forms zinc ions, and gives up electrons that account for the negative charge at this pole:

$$Zn \rightarrow Zn^{++} + 2e^{-}$$

The electrons move through an outside circuit to the carbon rod where they reduce the ammonium and manganese ions in the paste electrolyte. The products of the reactions at both poles diffuse into the paste, until eventually the reactants are used up and the cell is no longer useful.

Other combinations of oxidation-reduction reactions may serve as electrochemical cells, but their application to practical uses is limited by their cost and electrical capacity. A particular type of electrochemical cell that is attracting much interest because of its high efficiency and low cost is called a fuel cell. These cells require a continuous supply of raw

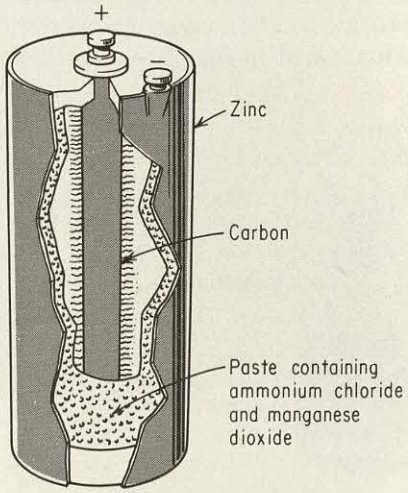


Fig. 19-5. The ordinary dry cell.

materials that are used up at the electrodes without changing the electrodes themselves. Favorite raw materials for the fuel cell include fuel gases such as natural gas and even powdered coal. One of the simpler fuel cells consumes hydrogen and oxygen which are "burned up" at porous carbon electrodes as shown in Fig. 19-6. The hydrogen undergoes oxidation at the left-hand electrode (−) and the oxygen undergoes reduction at the right-hand electrode (+). The electrolyte is an alkaline solution and the chemical product of this electrochemical cell is water. The efficiency of a fuel cell may be nearly twice that obtained from similar fuels when burned in the steam turbine or conventional internal combustion engine.

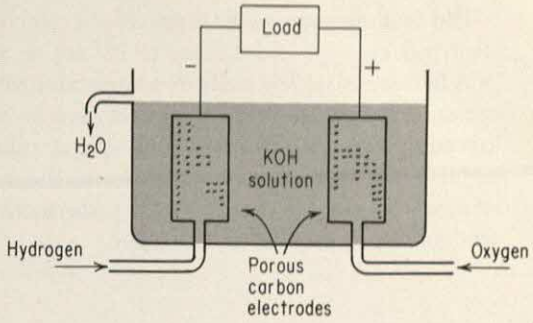


Fig. 19-6. A hydrogen-oxygen fuel cell.

The dry cell and the fuel cell are examples of *primary* cells, which are cells whose reactions are not readily reversible and consequently cannot be recharged. An electrochemical cell that can be recharged is called a *secondary* cell, and the most common secondary cell is the lead storage battery.

The common *storage battery* is a collection of at least three cells attached to each other in a series circuit. Each cell is capable of putting out slightly over 2 volts. A charged lead storage cell consists essentially of an electrode of lead (the negative pole), an electrode of lead dioxide (the positive pole), and an electrolyte of sulfuric acid (Fig. 19-7). When we use a lead storage battery, say, to start our car in the morning, we are making use of the energy difference that electrons have on these two electrodes. At the negative pole, lead goes into solution according to the reaction:

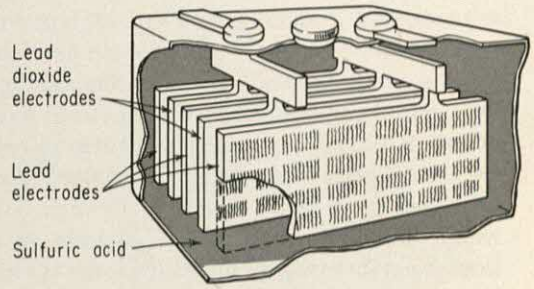


Fig. 19-7. The lead storage battery.

$$Pb \rightarrow Pb^{++} + 2e^{-}$$

At the positive pole, electrons are used up in the reaction:

$$PbO_2 + 4H^{+} + 2e^{-} \rightarrow Pb^{++} + 2H_2O$$

The hydrogen ions are supplied by the sulfuric acid. Both electrode reactions produce lead ions, but in the presence of sulfate ions from the sulfuric acid, lead sulfate is found at each electrode. Since lead sulfate, $PbSO_4$, is extremely insoluble, it remains mechanically attached to each electrode. The voltage of the lead cell arises from the fact that electrons on the lead possess 2 volts more energy than they do on the lead dioxide, and the resulting electrochemical potential accounts for the energy that may be obtained from the lead cell.

In recharging the lead cell, all the reactions that occur above are reversed by placing an outside voltage on the cell which forces electrons back in the opposite direction. The lead sulfate at the negative pole is reduced back to free lead while the lead sulfate at the positive pole is oxidized to PbO_2. The original sulfuric acid is regenerated, giving an electrolyte with a higher density than that of the electrolyte when the cell is discharged. In fact, a common method of examining the charge on a lead storage cell is to measure the density of the electrolyte.

The lead storage battery provides a convenient source of electrical energy. In addition to its use in automobiles, the lead battery is widely used in commercial trucks, telephones, mines, submarines, railway signals, and in emergency lighting equipment. We occasionally hear of other kinds of secondary cells that might be superior to the lead cell, but for economy, popularity, and general performance, the lead storage cell has never been exceeded.

19-3 Electrolysis

In Chapter 14 we saw how the early work of Faraday provided a firm basis for the further development of the concept of valence. In view of our modern notion of atomic structure, this concept is not difficult to appreciate, for valence is the manifestation of the electron gain or loss of an element, and Faraday's laws (p. 265) provided the quantitative relationship between valence and electrolysis. *Electrolysis*, you will recall, is the decomposition of a substance by electrical energy. That is, an outside source of electrical energy is placed on a chemical system to overcome or undo the effects that might occur in an electrochemical cell. For example, if we pass electricity through a solution of salt water, through melted sodium chloride, or through a solution of any electrolyte, the result is the decomposition of the system. At each electrode a certain chemical change occurs, provided that the voltage of the applied electricity exceeds the electrochemical potential of the substances that undergo chemical decomposition. Let us now look at the electrolysis of melted sodium chloride.

Melted sodium chloride consists of an equal number of sodium and chloride ions, which are free to move independently of each other. When we place two inert electrodes, such as graphite or platinum, into the melted sodium chloride, we find that an electric current flows through the melted salt quite easily (Fig. 19-8). In a short time, bubbles of chlorine gas will appear at the positive electrode, and at the negative electrode, a bright, shiny deposit of metallic sodium will appear. In order to explain the appearance of these two elements, let us digress a moment or two.

We are well acquainted with the intense chemical reactivity of sodium and chlorine. Sodium has a high positive electrochemical potential and hence undergoes oxidation readily. Chlorine, on the other hand, has a very low positive, or a high negative, electrochemical potential; its standard potential is -1.36 volts, compared to sodium's $+2.71$ volts. Thus, chlorine undergoes reduction, or electron gain, very easily. Since an electrochemical cell that provides for the simultaneous oxidation of sodium and the reduction of chlorine would, at standard conditions, yield 4.07 volts of electrical energy, in the electrolysis of sodium chloride we must supply electrical energy at a voltage in excess of that

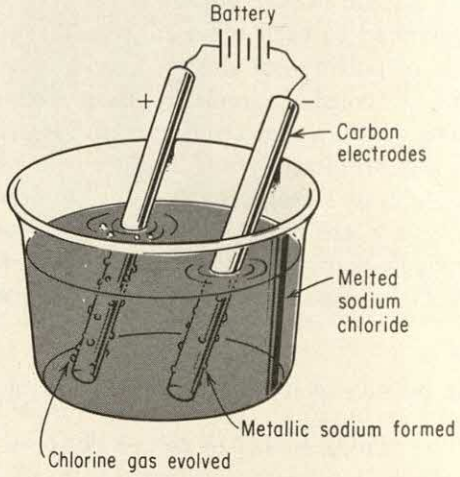


Fig. 19-8. The electrolysis of fused sodium chloride.

produced in the sodium-chlorine electrochemical cell. That is, we must reverse the oxidation-reduction process that ordinarily would occur spontaneously.

The mechanism of electrolysis of sodium chloride is essentially the migration of ions—cations to the cathode and anions to the anode—followed by reduction at one electrode and oxidation at the other. The sodium ions, attracted to the negative pole, pick up electrons and form sodium atoms:

$$Na^+ + e^- \rightarrow Na$$

Chloride ions migrate to the positive pole, lose electrons, and thus produce chlorine atoms:

$$Cl^- \rightarrow Cl + e^-$$

The sodium is reduced and the chlorine is oxidized, exactly the opposite of what would happen spontaneously. The only reason it can happen here is that we have supplied an outside source of energy in the form of an electrical potential that overcomes this natural tendency. Note that the melted sodium chloride does not conduct electrons through itself. It acts as an electrical conductor because the ions migrate independently and subsequently gain or lose electrons at the electrodes.

A solution of sodium chloride conducts electricity in a manner similar to that of the melted sodium chloride above. There is one important difference, however, and it becomes evident when we examine the products of the electrolysis process. A solution of sodium chloride contains, in addition to sodium ions and chloride ions, water molecules and a small quantity of hydrogen ions and hydroxyl ions, which result from the small but definite ionization of water. When suitable electrodes are placed in a salt solution, both the sodium ions and the hydrogen ions migrate toward the cathode (Fig. 19-9). But at the cathode only hydrogen is found; there is no evidence of any free sodium. With both sodium ions and hydrogen ions coming in contact with the cathode, we may expect the reactions:

$$Na^+ + e^- \rightarrow Na$$
$$H^+ + e^- \rightarrow H$$

Yet, only the hydrogen reaction occurs. From the list of electrochemical potentials, we see that sodium has a much higher solution tendency than does hydrogen. Conversely, hydrogen ions take up electrons and come out of solution much more readily than do sodium ions. Hence, only the hydrogen plates out at the cathode. Here we have an excellent example of an important principle in physical science that is often applied as well to other areas of knowledge. Although in the electrolysis of a sodium chloride solution two reactions may occur simultaneously at the cathode, the one that goes easier—the one that requires the lesser energy—is the one that takes place. Thus, in any natural process where there is a choice between two or more paths, the easiest path is the one taken.

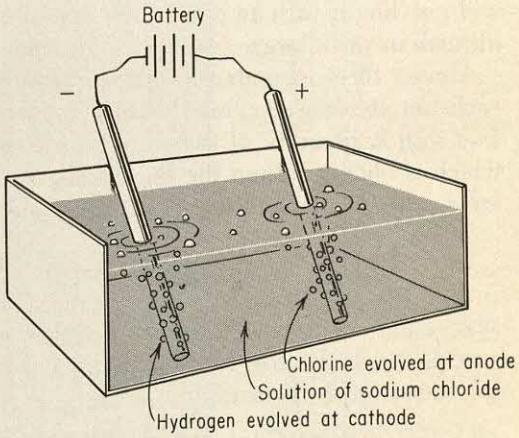


Fig. 19-9. The electrolysis of a solution of sodium chloride.

19-4 Electrolytic Production of Elements

We have seen how the electrolysis of melted sodium chloride produces sodium metal and chlorine gas, and electrolysis is essentially the industrial method of making both these elementary substances. Chlorine, however, is more commonly made by the electrolysis of a sodium chloride solution, primarily because hydrogen and sodium hydroxide are produced at the same time. Many other elements, such as fluorine, bromine, lithium, aluminum, calcium, and most of the magnesium used today, are produced by electrolysis. Other metals such as nickel, copper, and zinc are purified or recovered by electrolytic methods. Because the process for each metal is somewhat unique, we will not take time here to consider all of them in detail. However, a brief examination of the production of magnesium from sea water will provide us with an idea of the importance of electrolysis in modern metallurgy.

Almost three pounds of magnesium ions are present in each ton of sea water, in which the magnesium is dissolved as a salt. A quantity of sea water is first treated with lime, which is obtained from the huge banks of oyster shells that are readily available along certain seashores:

$$\underset{\text{lime}}{Ca(OH)_2} + \underset{\substack{\text{magnesium}\\\text{ion}}}{Mg^{++}} \rightarrow \underset{\substack{\text{calcium}\\\text{ion}}}{Ca^{++}} + \underset{\substack{\text{magnesium}\\\text{hydroxide}}}{Mg(OH)_2}$$

Magnesium hydroxide is quite insoluble, so it settles as a precipitate, and, after being separated from the sea water, it is dissolved in hydrochloric acid:

$$\underset{\substack{\text{hydrochloric}\\\text{acid}}}{Mg(OH)_2 + 2HCl} \rightarrow \underset{\substack{\text{magnesium}\\\text{chloride}}}{MgCl_2} + 2H_2O$$

The solution of magnesium chloride is then evaporated to the point where the solid magnesium chloride salt can be obtained. This salt is melted and then electrolyzed, giving magnesium metal and chlorine:

$$MgCl_2 \xrightarrow[\text{electrolysis}]{} Mg + Cl_2$$

Precisely the same mechanism of electrolysis is involved here as in the electrolysis of melted sodium chloride. The by-product, chlorine, is used in an interesting fashion to produce hydrochloric acid. This is accomplished by burning natural gas, primarily methane, in a chlorine atmosphere:

$$\underset{\text{methane}}{CH_4} + 2Cl_2 \rightarrow 4HCl + C$$

The hydrochloric acid is recovered and used to dissolve magnesium hydroxide in an earlier step in the process. The carbon that is formed here is of the finely divided kind sometimes called *carbon black* or *lampblack*. It finds wide use as a pigment and as a component of ordinary black rubber. A magnesium plant using this method must be located near ample supplies of calcium carbonate (or oyster shells),

natural gas, and cheap electrical power. Such a situation exists along the Gulf Coast of Texas, where successful plants of this type have been in operation since the advent of World War II.

Another important application of electrolysis is in the electroplating of metals. Since most of us prefer to see a bright, shiny surface on metallic objects, much of our jewelry, the ornaments on our automobiles, and many useful industrial tools are electroplated. The plating of silver is one of the most common applications of electrolysis in this field. Silver-plated tableware is first fashioned of copper or brass, then placed as a cathode in an electrolysis system. A piece of pure silver serves as the anode, and the electrolyte is a solution of silver cyanide. As electrical current passes through the electrolysis system, silver dissolves at the anode and plates out at the cathode. The temperature, amperage, and concentration of the electrolyte are carefully controlled in order to achieve a uniform, fine-grained deposit of silver on the object. After buffing and polishing, the silver-plated object may appear as attractive as if it were made of solid silver.

A widely used method of studying the structure of subsurface rocks and minerals relies upon electrolysis. This method is known as *electric well logging*. When an oil well is being drilled, electrodes are placed in the well along with the drill. An electrical voltage is placed on the two electrodes, and the amount of current that passes between the electrodes is a measure of the quantity and type of electrolyte between the electrodes. In sandstone formations, only a small amount of current flows, but in a limestone or shale zone there are more soluble electrolytes and considerably more current passes through the electrolysis system. A record of this information kept as the well is being drilled (Fig. 19-10) gives valuable information about the rocks, their porosity, and the nature of the liquid that fills the rock pores.

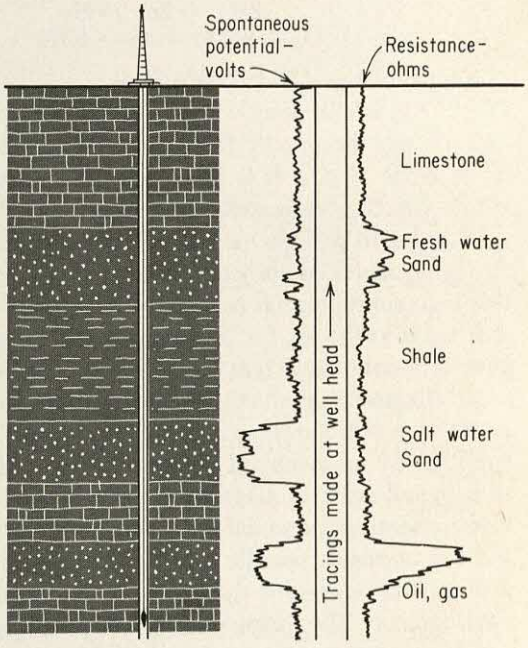


Fig. 19-10. A geological structure studied by an electric well log.

19-5 Corrosion

The natural tendency of metals to react chemically with their environment is called *corrosion*. When the environment consists of air, moisture, and possibly some electrolytes, all the essentials are present for an electrochemical cell. Furthermore, the corrosion tendencies of metals are closely related to their electrochemical potentials. Even the less active metals, such as copper and silver, undergo corrosion. However, we will concern ourselves mostly with the corrosion of iron because this metal is vital to our modern culture and at the same time is very susceptible to corrosion. The tremendous expenses incurred to prevent or halt the corrosion of iron is indicative of the importance of this problem.

Iron corrodes, according to our best knowledge, by an

electrochemical process. Impurities or stresses in the metal cause certain areas to show greater solution tendencies than others, and these areas become anodes, where oxidation occurs (Fig. 19-11). Here the iron goes into solution as the ferrous ion with a loss of two electrons:

$$\underset{\text{ferrous ion}}{Fe \rightarrow Fe^{++}} + 2e^-$$

The less reactive areas become cathodes, where reduction occurs. Either or both of two simple reduction processes may occur, depending upon the acidity and the presence of oxygen and other substances:

$$2H^+ + 2e^- \rightarrow H_2$$
$$O_2 + 2H_2O + 4e \rightarrow 4OH^-$$

The ferrous ions (Fe^{++}) are further oxidized to ferric ions (Fe^{+++}), which react with the hydroxyl ions formed at the cathode and eventually form ferric hydroxide, $Fe(OH)_3$, or ferric oxide Fe_2O_3. It is these last two substances that constitute what we commonly call *rust*.

In order to prevent or control corrosion, we must inhibit the electrochemical process. A very direct method of doing this is to superpose an outside voltage upon the iron so that it is made cathodic, for it is only at the anode that the iron goes into solution. Thus, pipelines, bridges, and ships have been effectively protected from corrosion by making their entire iron structures cathodic. The outside voltage may be supplied by an electrical generator or by attaching the iron to a metal, such as magnesium and zinc, that has a higher electrochemical potential than iron. Other methods of controlling corrosion usually involve a barrier such as paint or a thin plate of another metal which isolates the iron from its environment. The paint usually contains pigments such as red lead or zinc chromate that inhibit corrosion.

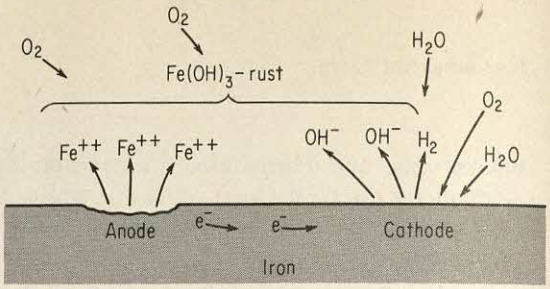


Fig. 19-11. The electrochemical mechanism of corrosion of iron.

19-6 Semiconductors

Up to now we have become acquainted with two kinds of electrical conductors—metals and electrolytes. Metals, we found, conduct electricity by merely passing electrons along from atom to atom through their solid structure. But electricity passes through a solution of an electrolyte by the movement of ions and the chemical change of these ions at the electrodes. A third kind of conductor, sometimes called a *semiconductor*, conducts electricity by a mechanism entirely different than that of either metals or electrolytes.

Metals are characterized by the presence of many freely moving electrons, whereas good insulators have virtually no electrons that can be easily moved—all their outer electrons are tied up in covalent bonds. The gradation between metallic elements and nonmetallic elements on the periodic table (p. 254) appears in those elements that are intermediate between conductors and insulators. These are the ele-

ments in Group IVB: carbon, silicon, and germanium. All of them may exist in the diamond structure in which each atom shares its four valence electrons with four other atoms. Thus, all the available electrons of each atom are tied up in four covalent bonds.

Pure germanium, in the diamond structure, is ordinarily a good insulator, but when very small amounts of impurities —less than one atom in one hundred million—are purposely added to it, it becomes a semiconductor. Suppose we add to it a small amount of an element that has five valence electrons, such as phosphorus, arsenic, or antimony. Each time one of these atoms enters the structure there will be an extra electron that is not tied up in covalent bonds (Fig. 19-12).

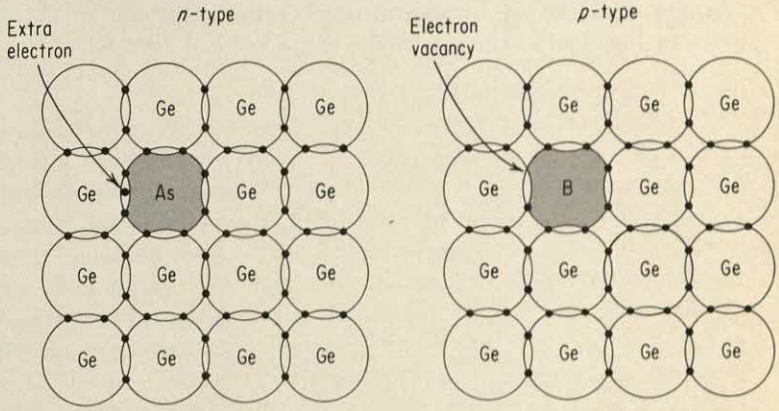


Fig. 19-12. The structure of a semiconductor.

As a result, this system has enough free electrons in its structure to make it a conductor.

Suppose we now add to the germanium an impurity that has only three valence electrons. Each time an atom of such an impurity as boron, aluminum, or gallium enters the structure, there will be a shortage of one electron and thus a vacancy for an electron at each of the substituted atoms. If we place a small electrical potential on this system, electrons will migrate to these vacancies, creating new vacancies at the same time. This process of "give-and-take" amounts to electrical conduction of a sort and is called conduction by "holes." This kind of semiconductor with a deficiency of electrons is called a p-type (positive), and one having excess electrons is called an n-type (negative).

When n-type and p-type semiconductors are placed next to each other, a very powerful and efficient means of changing alternating to direct current is created. For instance, in Fig. 19-13a, two semiconductors are joined intimately in an electrical circuit. When a voltage is applied to the p and n pair, electrons in the n-type material move from right to left toward the p-n boundary at the same time that holes move from left to right toward the boundary. If the voltage is applied in the opposite direction, as shown in Fig. 19-13b, the electrons and the holes move away from the boundary, but practically no current passes through the junction. Conse-

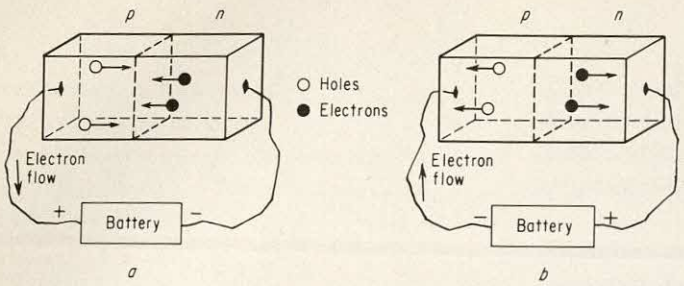


Fig. 19-13. A simple combination of a p- and an n-type semiconductor.

quently, such a system serves as a valve that permits electrons to flow only in one direction, and in this respect it can be used to convert alternating current to direct current. This is called rectification.

A more interesting arrangement of semiconductors is shown in Fig. 19-14, where a p-type is sandwiched between

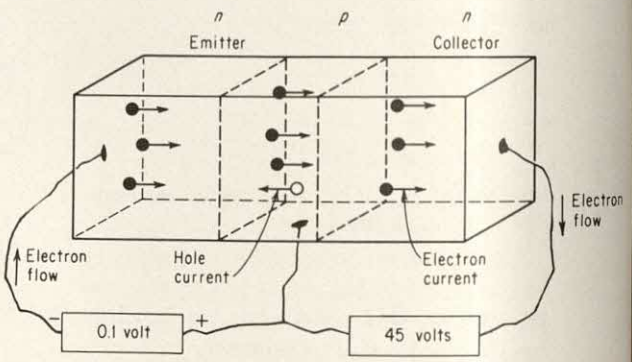


Fig. 19-14. A multiple combination of p- and n-type semiconductors, commonly called a transistor.

two n-types. In this arrangement, the p-type semiconductor is extremely thin—only a thousandth of an inch thick—and, compared to its n-type neighbors, it is sparsely populated with holes. The rule that electrons can flow in the n to p direction at a junction, and not in the opposite direction is now modified by virtue of the presence of a third component in the system. For example, at the junction on the left—called the *emitter junction*—electrons can readily move across the boundary from left to right, and a relatively smaller number of holes move in the opposite direction. The electrons crossing the emitter junction traverse the thin section of the p-type semiconductor and readily cross over the collector junction into the n-type material on the right. However, if no electrons passed through the emitter junction, none would flow through the collector junction. Accordingly, as the diagram shows, a low voltage across the emitter junction controls a much higher voltage across the collector junction. Although the details of the theory of this system are beyond our present interest, we can see that a large amount of power can be controlled by a small amount of input power. This control of power, permitting amplification and rectification, is essential to all kinds of electronic devices

such as radio, television, and telephone, and the advantage of semiconductors over the conventional vacuum tube is based on the fact that semiconductors require but one-millionth the power input needed by a vacuum tube, and the fact that semiconductors are extremely long-lived compared to vacuum tubes.

19-7 Summary

All chemical reactions involve electrical energy, a manifestation of the fact that all elements tend to either gain or lose electrons. This tendency is measured in terms of electrochemical potential.

The electrochemical potentials of many chemical changes have been measured, and a few of these have been adapted to useful purposes in such devices as the common dry cell and the lead storage battery.

The conduction of electrical current by a solution always involves electrolysis, which is essentially the movement of ions. Oxidation occurs at one electrode and reduction at the other. Electrolysis of melted sodium chloride produces free sodium and chlorine gas. Other elements, including magnesium, are also industrially produced by electrolysis.

Electroplating is an application of electrolysis in which one metal is plated onto an electrode for the protection or decoration of the base metal.

Corrosion is an electrochemical process in which anode and cathode areas account for the metal being attacked by its environment.

Semiconductors, usually made of Group IVB elements that contain small but controlled amounts of impurities, act as n-type (excess of electrons) or p-type (shortage of electrons) conductors. Combinations of these materials make efficient rectifiers and amplifiers that can be used in small radios and other communication instruments.

EXERCISES

A. KEY TERMS TO REMEMBER

anode
cathode
corrosion
dry cell
electrochemical cell
electrode
electrolysis
fuel cell
primary cell
secondary cell
semiconductor
standard electrochemical potential
storage battery

B. QUESTIONS ABOUT ELECTROCHEMICAL ENERGY

1. From the data in Table 19-1, which two metals would you select to make an electrochemical cell with a maximum voltage?

2. Show by a diagram how you might construct the cell in the preceding question.

3. In general, what kind of chemical reaction is used in an electrochemical cell?

4. By a diagram show how you would attach a battery charger to a lead storage battery in order to recharge it. Show the reactions that take place at each electrode when the battery is being recharged.

5. In the electrolysis of a sodium chloride solution, only hydrogen forms at the cathode. At the anode either chlorine or oxygen could form. Which of these two actually is formed, and why?

6. List the raw materials that are required to obtain magnesium from sea water. Show how electrolysis is an important step in this process.

7. Assuming that ideal conditions prevail, which process would require more electricity—the production, by electrolysis, of a pound of aluminum or a pound of sodium? Assume that there are approximately the same number of atoms of each metal in both cases.

8. In plating silver on silverware, do you need a particular voltage? Is there anything unique about the amperage required?

9. If in making an electric log of a well, you encountered a region that showed excessively high electrical conductivity, what would you conclude?

10. In terms of electrochemical potential and the possibility of corrosion, comment on the advisability of the following home-building practices:

a. Nailing a galvanized (zinc coated) roof with iron nails.
b. Nailing a copper roof with iron nails.
c. Attaching an iron pipe to a copper pipe.
d. Using a brass valve on a galvanized pipe.
e. Grounding the telephone (a direct current instrument) to a water pipe.

11. Tell how a solid substance can conduct electricity by having either an excess or a shortage of electrons in its structure.

12. What are some of the advantages of a semiconductor type of amplifier in a radio compared to the traditional vacuum tube amplifier, other than their difference in size?

C. PROBLEMS ABOUT ELECTROCHEMICAL ENERGY

1. What is the voltage of a standard cell constructed with iron and copper and involving reactions given in Table 19-1?

2. What would be the highest voltage obtainable from the reactions listed in Table 19-1? Ans. 4.60 volts. Is such a cell practical?

3. One type of hearing aid battery employs zinc for one electrode and mercury for the other. Approximately what voltage would you expect from such a cell, assuming that the standard electrochemical potentials would provide a clue?

4. Tin is more electronegative than iron, and zinc is more electropositive than iron. Explain why both of these metals are used to protect iron (tinplate and galvanize), even though the electrochemical effects are diametrically opposite to each other.

Twenty

Heat Energy

In the previous chapters we have studied a number of different kinds of energy—chemical energy, electrical energy, kinetic, and potential energy. The form of energy that we shall consider in this chapter—heat energy—is even more closely associated with life and living things than these other forms of energy. Man has produced temperatures on earth as low as —459°F and as high as 180,000,000°F (Fig. 20-1), yet in order to survive, his body temperature must not drop much below 96°F nor rise much above 102°F. Man's body, then, must be a carefully controlled heat engine with very special temperature-regulating mechanisms. Only through the discovery of how to make fire was man able to control his environment sufficiently to be able to move out of the semitropical regions and, eventually, to conquer the planet. The development of heat engines has helped release man from drudgery and has given him sufficient time to be inventive and speculative, to become inventor, economist, poet, philosopher; and scientist.

The late Dr. Karl T. Compton, president of the Massachusetts Institute of Technology, has said, "In the last fifty years physics has exerted a more powerful beneficial influence on the intellectual, economic and social life of the world than has been exerted in a comparable time by any other agency in history. Its influence has far exceeded that of wars, political alignment or social theories." Lloyd W. Taylor, in commenting on Dr. Compton's remarks, added that "President Compton was unduly restrained. His fifty-year span should have been two hundred. The development of steam power, intimately associated with the beginning of scientific comprehension of the nature of heat, took place during the

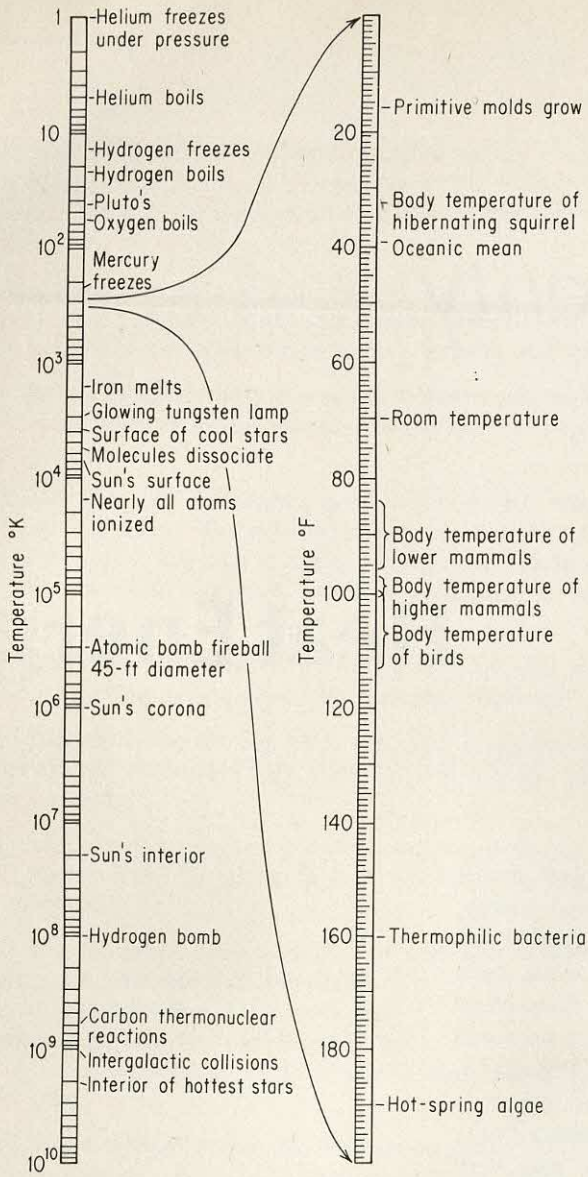


Fig. 20-1. Temperatures from the lowest to the highest known to man, and the narrow range of temperature within which life exists.

eighteenth century. Out of it grew industrial society. . . . Out of the early study of heat grew also that broadest of all scientific laws, the doctrine of Conservation of Energy. Whether viewed from the utilitarian or the purely scientific standpoint, the beginning of our comprehension of the nature of heat marks an epoch in human living and thinking." *

The understanding of heat energy, like most other understandings, did not come to man easily. One of the first ways man used heat energy (possibly as early as 10,000 B.C.) was to bake building materials such as bricks. By 6000 B.C., petroleum and asphalt were important items of commerce and are thought by some to have been used for heating and for the refining of ore. Except for a few toys developed by

* Lloyd W. Taylor, *Physics, The Pioneer Science*, reprinted through permission by Dover Publications, Inc., New York, p. 248.

the Egyptians, man did not have enough knowledge about heat energy to develop an engine to do his bidding until James Watt put steam to work in 1769. Up until this time, man used heat primarily to keep himself warm and to provide light so he could control, in a primitive fashion, his environment.

20-1 The Measurement of Temperature

The first advance beyond this primitive use of heat energy was largely dependent upon an accurate measuring instrument for heat. Lord Kelvin's statement is again relevant: "When you can measure what you are speaking about and express it in numbers, you know something about it." But it was not until the time of Galileo, who is given credit for developing the first "thermometer" in A.D. 1600, that the way was opened for a more complete understanding of heat. From Fig. 20-2 it can be seen that Galileo's thermometer consisted essentially of an enclosed gas that changed volume with changing temperatures. As the thermometer was heated, the gas in the bulb expanded and drove the liquid lower in the vertical tube. When the thermometer was cooled, the gas contracted, and the air pressure forced the liquid to rise. Although this device was quite sensitive to temperature differences, it was also affected by differences in the atmospheric pressure, and it was not until some years later that scientists recognized this problem and solved it by placing the bulb of the thermometer at the bottom. In this position, changes in atmospheric pressure can affect only the liquid, which is relatively incompressible in comparison to the air in Galileo's thermometer.

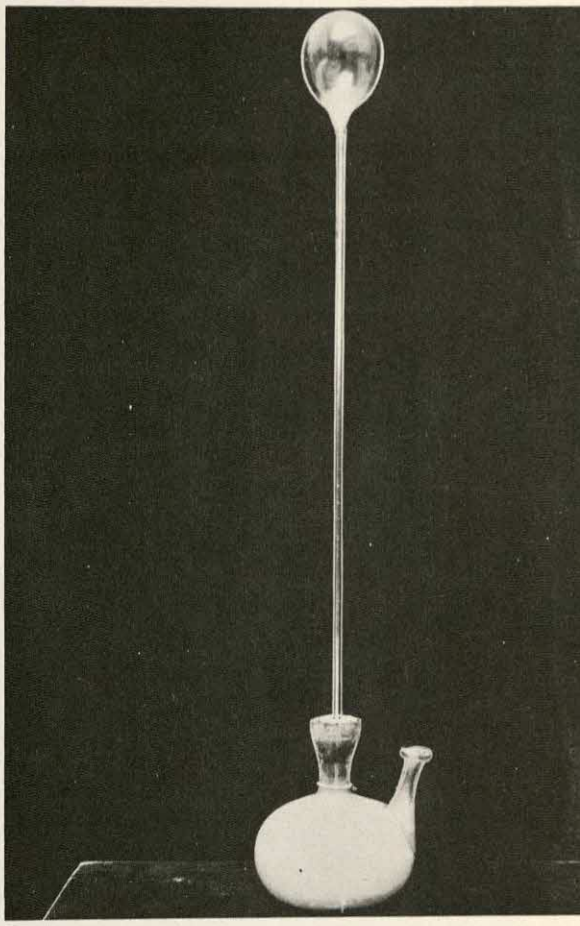


Fig. 20-2. Photograph of a surviving model of Galileo's thermoscope. (Courtesy Fisher Scientific Company)

Increased usefulness and accuracy in a thermometer can be gained if a liquid is used which has a low freezing point and a high boiling point, and which expands uniformly with changes in temperature. Scientists soon discovered that mercury scores high on all three of these counts and is therefore an excellent thermometric liquid. The next advance in temperature measurement came in the late seventeenth century when the need for several points of reference or "fixed points" on a temperature scale became apparent. These "fixed points" are temperatures that are presumed to be the same any place on earth as long as the pressure does not change—for instance, the melting point of ice is dependable enough to be used as a point of reference for measuring other temperatures—and this use of reference points allowed workers at different places to compare temperatures accurately. They could say that the temperature was, for instance, one-third of the way from the temperature of a salt-ice mixture to the temperature of the human body, and, since this was a ratio that did not depend upon the exact dimensions of the instrument, this method greatly extended the usefulness of any thermometer.

STATE INSTITUTE OF EDUCATION Govt. of West Bengal

Newton observed that a mixture of snow and water always gave the same reading on his thermometer. Guillaume Amontons, a French physicist, found that water always boiled at the same temperature. From these two observations, the Swedish astronomer, Anders Celsius, in 1742 conceived the idea of using these two temperatures, the melting point and the boiling point of water, as the fixed points of a temperature scale. Dividing the intervening distance on his thermometer into 100 parts or degrees, he called the freezing point 100° and the boiling point 0°. The following year this temperature scale was reversed so that the 0° represented the freezing point and the 100° the boiling point. This modified form is now known as the *Celsius thermometric scale.* The Celsius scale (formerly known as the centigrade scale) is now used universally by scientists as well as by most nations. The relationship between the Celsius and Fahrenheit temperature scales is illustrated in Fig. 20-3.

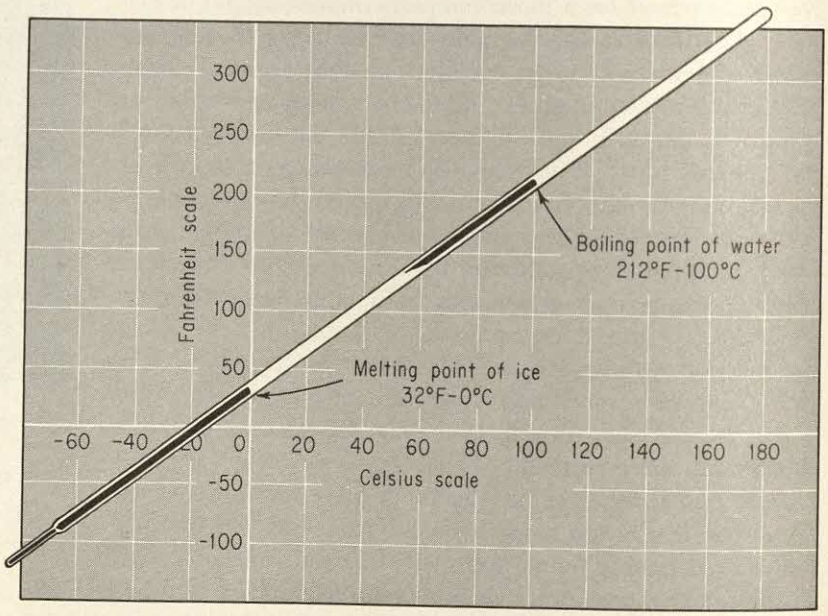


Fig. 20-3. A comparison of thermometric scales—Celsius and Fahrenheit.

It may seem incredible that it took 142 years to develop the thermometer, but we must remember that new ideas come slowly even now and that the methods of science were certainly not well developed in the year 1600.

20-2 Heat and Temperature

Temperature and heat are often confused with each other. *Temperature* is a measure of the *average* kinetic energy of the molecules of a substance; the higher the temperature the faster the molecules move. *Heat,* however, is a measure of the *total* kinetic energy of the molecules in a system. For instance, a cup of hot coffee has a higher temperature than the ocean because the molecules in the cup of coffee have

a higher kinetic energy than the molecules in the ocean. But, regardless of the temperature of the cup of water, the ocean contains a greater amount of heat since it has so many more molecules and therefore greater total kinetic energy.

In order to undertsand heat energy we must know the units in which it is measured. Heat is measured in calories or in Btu's (British thermal units). A *calorie* is the amount of heat required to raise the temperature of one gram of water one degree Celsius. A *Btu* is the amount of heat required to raise the temperature of a pound of water one degree Fahrenheit. The Btu is a considerably larger unit of heat than the calorie; one Btu equals 252 calories.

20-3 Specific Heat

Different materials require different amounts of heat to cause a given change in temperature. The sun on a warm, summer day, for instance, raises the temperature of the sand on a beach more rapidly than that of the water in a lake. The simple experiment diagrammed in Fig. 20-4 serves to illustrate this phenomenon. Suppose we put 100 g of dry sand in a double boiler above boiling water until it is the same temperature as the steam. Now let us quickly pour this hot (100°C) dry sand into 100 g of water, the temperature of which is 20°C. As the sand and water are stirred with a thermometer, the temperature of the mixture becomes 35°C. When the 100 g of sand cools from 100°C to 35°C it warms the 100 g of water from 20°C to 35°C; the water, in other words, gains 100 g $\times$ 15°C $\times$ 1 cal/g-°C, or 1500 calories. Since energy can be neither created nor destroyed, this 1500 calories of heat energy must have been transferred from the sand to the water, assuming no heat loss or heat gain from the environment. We know that the sand gave up 1500 calories of heat energy when 100 g of it cooled from 100°C to 35°C, and therefore each gram of sand must have given up 0.23 calories for each degree it cooled:

$$\frac{1500 \text{ cal}}{100 \text{ g} \times 65°\text{C}} = 0.23 \text{ cal/g-°C}$$

For our purpose, we shall call the amount of heat needed to change 1 g of a substance 1°C its *specific* heat. From the definition of the calorie and the Btu, it is apparent that the specific heat has the same numerical value whether we measure it in cal/g-°C or in Btu/lb-°F. In our problem then, the sand has a specific heat of 0.23 cal/g-°C.

Stated another way, we could say that sand has 23/100 of the heat capacity of an equal mass of water.

We now have a way of determining the amount of heat that is associated with any change in temperature that does not involve a chemical change, a transformation from a solid to a liquid, or a transformation from a liquid to a gas. If we multiply the mass by the change in temperature and the

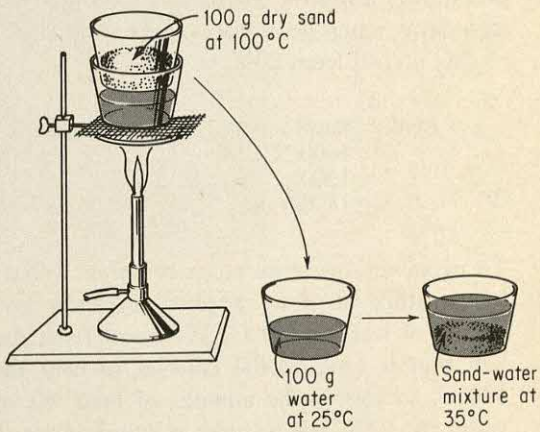


Fig. 20-4. A simple experiment by which the heat capacity of sand can be determined.

specific heat, we will have the heat expressed in Btu or cal, depending upon the units we use:

$$\text{heat} = \text{mass} \times \text{change in temperature} \times \text{specific heat} \quad (20.1)$$

Table 20-1 lists the specific heats of some common substances. An illustration or two will make the use of formula 20.1 clear. Suppose we have a 1500-g frying pan made of iron. How much heat will be needed to raise the temperature of this utensil from 20°C to 100°C?

$$\begin{aligned}\text{heat} &= \text{mass} \times \text{change in temperature} \times \text{specific heat} \\ &= 1500 \times (100 - 20)\,^{\circ}\text{C} \times 0.12 \text{ cal/g-}^{\circ}\text{C} \\ &= 1500 \times 80 \times 0.12 \\ &= 14{,}400 \text{ cal}\end{aligned}$$

To make the question more complex: What would the final temperature be if we applied the same amount of heat to 1500 g of water at 20°C? We know from the previous problem that it took 14,400 calories to heat the frying pan to 100°C, so this is the amount of heat we will apply to the water. At 20°C the water is a liquid so we shall use the specific heat of water as a liquid and substitute that value in formula 20.1, letting X equal the change in temperature.

$$\begin{aligned}\text{heat} &= \text{mass} \times \text{change in temperature} \times \text{specific heat} \\ 14{,}400 \text{ cal} &= 1500 \text{ g} \times X\,^{\circ}\text{C} \times 1 \text{ cal/g-}^{\circ}\text{C}\end{aligned}$$

$$\frac{14{,}400}{1500 \times 1} = X\,^{\circ}\text{C}$$

$$9.6\,^{\circ}\text{C} = X$$

The change in temperature would be 9.6°C. Since heat was being added, this would be a rise in temperature, and the final temperature would be 9.6°C + 20°C, or 29.6°C. We can see from this example that it takes much more heat to change the temperature of water by a given amount than it does to change the temperature of iron by an equivalent amount. The difference is due to the fact that water has a much higher specific heat than iron.

When we examine Table 20-1, we see that the specific heat of water is very high compared to that of nearly every other common substance. In fact, it is only rarely that we encounter a substance that has a greater specific heat than water. Water therefore is a good carrier of heat energy and also an excellent cooling agent. Hot-water heating systems, hot-water bottles, and automobile radiator systems all depend upon the high heat capacity of water for their effectiveness. Ocean currents moving from tropical areas to cooler regions give off enormous quantities of heat as they cool a few degrees. It is said that an ocean current such as the Gulf Stream delivers heat at the rate of 9.2×10^{14} Btu/hr as it cools 1°F; this amount of heat is equivalent to that given

TABLE 20-1 SPECIFIC HEATS OF SOME COMMON SUBSTANCES

(Cal/g-°C or Btu/lb-°F)

	Solid	*Liquid*	*Gas*
Aluminum	.217	.25	
Ammonia	.50	1.1	.52
Basalt	.20		
Copper	.093	.10	
Ethyl alcohol	.23	.55	.4
Ethylene glycol	.53	.57	
Granite	.19		
Iron	.12		
Mercury	.03	.033	.02
Oxygen	.34	.39	.22
Water	.5	1.00	.5

off by the burning of 40,000,000 tons of coal. The specific heat of water is so much greater than that of air that 1 cu ft of water in cooling 1°C will supply enough heat to warm 3000 cu ft of air 1°C. Thus, air near large bodies of water may have its temperature modified greatly. As the Gulf Stream flows from the tropical regions northward along the eastern seaboard (Fig. 20-5), it warms and moistens large

Fig. 20-5. The Gulf Stream and other major surface currents of the Atlantic Ocean. (From Sverdrup, Fleming, and Johnson. *The Oceans*, Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1942)

volumes of air, and, in transferring these enormous quantities of heat, it modifies the weather of our eastern states and is a dominant factor in the climate of western Europe.

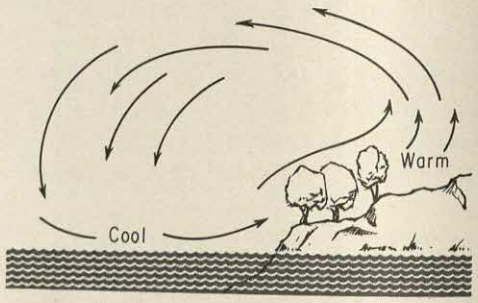
A uniform body temperature is necessary for our health and, indeed, for our survival. Our bodies provide the envi-

ronment in which the chemical reactions necessary to life take place. Since almost all chemical changes take place more rapidly as the temperature is increased, it is important that the temperature of our bodies be held nearly constant. As we saw in Chapter 17, the reaction velocity usually doubles for each 10°C rise in temperature, and, for this reason, if our body temperature changes much, it drastically affects our body chemistry.

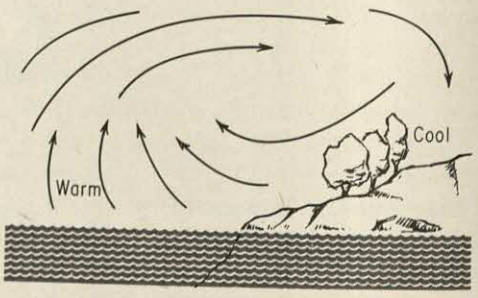
In medicine, use is being made of lower body temperatures in certain surgical operations. As the body temperature drops, the chemical activity decreases, and since with the slower chemical reaction rate the body needs smaller amounts of oxygen, it may be possible to perform operations that temporarily interrupt the flow of blood, which carries life-giving oxygen to all parts of the body. The time available for operating may consequently be increased, and some operations that cannot possibly succeed at normal body temperatures may now be performed with relative safety.

Since our bodies are composed of about two-thirds water, a large amount of heat must be exchanged in order to cause our body temperature to rise or fall very much. This fact makes it possible for us to go outdoors without freezing when the temperature is very low; the high specific heat of water enables us to lose large quantities of heat without much change in our body temperature. In an average day we eat food that has a fuel value of approximately 3,000,000 calories (or 3000 large calories),* which is enough energy to raise the temperature of a 150-lb man by more than 80°F. If our bodies were made of alcohol instead of water, this potential temperature rise would be nearly 160°F. To maintain a uniform temperature, our bodies must give off large quantities of heat. We have all experienced the proof of this when we have been in a room, crowded with people, in which the temperature has risen noticeably. The food we eat, in addition to maintaining our body temperature, gives us the energy to do our work and to keep our bodies in repair, and these processes too depend upon the control of our body temperatures, whether it be through the evaporation of perspiration, control of the flow of our blood supply, or the high specific heat of water.

We have seen that the specific heat of rock and sand is low compared to that of water. As a result, the sun can heat an area of land to a much higher temperature than an equivalent area of water. Land areas also cool more rapidly than water. This fact is responsible for warm air often being produced over the land while cool air is produced over the water, causing cool air to blow inland during the day, and the cooler "land" air to blow seaward at night when the land cools, as shown in Fig. 20-6.



Sea breeze produced when land is warmer than adjacent water



Land breeze produced when land is cooler than adjacent water

Fig. 20-6. The movements of air masses at a shore line.

* A large calorie, often called a kilogram calorie, is 1000 times as large as the small calorie. The large calorie is used most commonly in connection with foods and diets.

20-4 Heat Energy Transformed

In Section 17-1 we already touched on many of the forms energy may take, and later in Chapter 28 we shall consider energy conversions from a different point of view. For the moment, however, let us consider only the change of heat energy into mechanical work. It was toward the end of the eighteenth century when another great leap forward was made in our comprehension of heat energy. Prior to this time, no one had realized that work could produce heat or that heat energy could be made to do work. Man had had to rely on slaves or animal labor to do his work, but when he found that heat energy could labor for him, the machine age was ushered in, with its increased standard of living and great technical achievements, all of which started with a very simple observation of the relationship between heat and work.

In the year 1789, Benjamin Thompson (later designated Count Rumford), a native of Rumford, New Hampshire, made a profound discovery while he was supervising the boring of a cannon. When a blunt drill was used one day, little happened except that an enormous amount of heat was created. Since he could produce almost as much heat as he wished, Count Rumford reasoned that this heat must come from the mechanical energy supplied by the engine that drove the boring tool. His was a historic discovery, for it was the first time anyone had observed a connection between work and heat. It remained, however, for Joule in England to conduct the experiments that showed the exact equivalence of energy and heat.

It takes a great amount of work to produce even a small amount of heat. To produce one Btu, for example, 778 ft-lb of work must be converted into heat energy. This means that the temperature of a pound of water, falling 778 feet and striking the ground, will increase 1°F if all its kinetic energy is converted into heat energy. In other words: 1 Btu is equivalent to 778 ft-lb of work.

The amount of work necessary to heat water is almost unbelievable. For instance, in order to bring only 2 pounds of water (about a quart) to the boiling point from 62°F, 300 Btu, or 300 Btu $\times$ 778 ft-lb/Btu, or 233,400 ft-lb of work are required. This is enough energy to lift one ton a distance of 116 feet, and if it were done in one minute, it would require 7 horsepower. If, to use another example, the Gulf Stream delivers 9.2×10^{14} Btu/hr as it cools just 1°F, it gives off energy equivalent to more than 7×10^{17} ft-lb/hr—or 360 billion horsepower. Over the earth as a whole, this equivalence of heat and energy means that enormous amounts of energy are carried in the form of heat by air and ocean currents.

You can now see why we get so much work done by a relatively small amount of fuel. When a pound of gasoline burns, it produces 20,000 Btu and therefore is capable of

doing 15,560,000 ft-lb of work for us. A car, then, can go hundreds of miles on a tank of gasoline; and a diesel locomotive, by using the energy of one teaspoonful of diesel oil, can move one ton for a mile. Man, through the control and use of heat energy, has greatly increased the amount of work he can get done in a day. Only by utilizing the tremendous storehouse of energy in our fossil fuels—coal and petroleum —have we been able to do the immense amounts of work necessary to maintain our civilization. Probably without realizing it, most of us have lived through one fuel revolution and will probably live through another. In 1920, 72 per cent of our energy needs were met by the use of coal, while only 10 per cent was obtained from petroleum and natural gas. By 1940 the percentages had changed greatly—coal, 52 per cent; oil and gas, 44 per cent—and before 1950 the United States was obtaining more energy from oil and gas than from coal. This change from a solid fuel economy to one based on liquid fuels has been so rapid that over half of all the oil and gas that has ever been consumed has been used since 1940.

We do not know exactly how large a reserve of fossil fuels remains in the earth or how long these fuels will last. It is generally agreed, however, that the peak production of oil and natural gas in this country may be reached by 1975, whereas for coal the probable date is 1990. Long before our fossil fuels are in short supply, we must make use of atomic power and develop research into the better utilization of solar energy and the harnessing of thermonuclear energy. If man is successful in this realm of scientific endeavor he will have made one of his greatest advances, for the power needs of man have grown so rapidly that only through this kind of imaginative development can civilization as we know it hope to survive.

20-5 Expansion Due to Heat

Nearly all materials, whether solids, liquids, or gases, expand as their temperatures increase. Gases expand more than liquids and liquids expand more than solids when they are heated. Precautions must be taken in many construction projects to allow for this expansion and for the corresponding contraction that occurs when the material cools. At the ends of large bridges, for instance, it is common to find sliding plates on the roadway and rollers or some other sliding mechanism under the main structure. Concrete highways and railroad tracks must have special joints to absorb the expansion and contraction of the concrete and rails. You may be interested to know that the side of a building that faces the sun expands and causes the building to lean slightly away from the sun. As an example, the expansion of the hot side of the Eiffel Tower makes it lean as much as six inches on a warm day.

Let us consider two basic examples of the expansion, or volume change, that accompanies an increase in temperature. One of these examples concerns the expansion of a gas, the other the behavior of water under changing temperatures. At 0°C the volume of any gas will increase or decrease by $\frac{1}{273}$ of its volume for each degree of temperature change (Celsius). Suppose we carry out an experiment to clarify this relationship. If we start with two different quantities of gas and record their volumes as their temperatures increase, the typical data we might collect are given in Table 20-2.

TABLE 20-2

Trial 1		*Trial 2*	
Volume	*Temperature*	*Volume*	*Temperature*
50 cm^3	−136.5°C	50 cm^3	−204.75°C
100 cm^3	0°C	200 cm^3	0°C
200 cm^3	273°C	400 cm^3	273°C

When we plot these data on a graph (Fig. 20-7), we find that the two lines drawn from these data meet at −273°C.

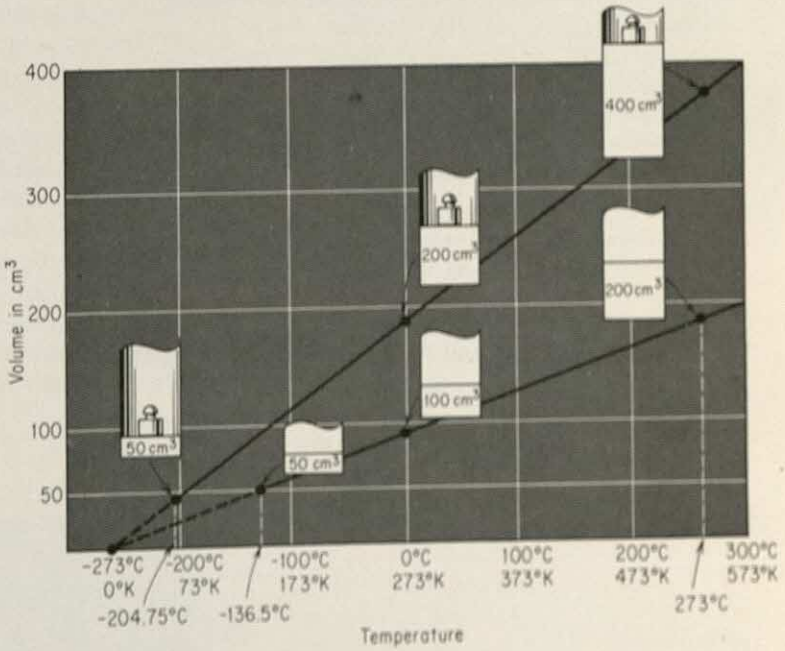


Fig. 20-7. An experiment relating the volume to the temperature of a gas. The experimental results are superimposed on a graph showing how 0°K is found.

The graph in Fig. 20-7 raises an interesting question. What would happen to the volume of a gas if the temperature were reduced to −273°C or 0°K? Since all known gases liquefy before they drop to such a low temperature, the question is purely academic, but theoretically, the volume of a gas would become zero at that temperature. The significance of such a starting point for a thermometric scale therefore becomes evident, and we call this temperature *absolute zero* (−273°C).

By starting a new scale at −273°C, we will produce a "gas" temperature scale, called a *Kelvin* (or *absolute*) *temperature scale* after Lord Kelvin, the British scientist who first proposed it. This scale uses the same magnitude of temperature variation per degree as the Celsius scale, but starts with a zero that is 273° lower. In other words:

$$°K = °C + 273 \qquad (20.2)$$

Now that we have an absolute temperature scale, we may state the relationship between gas volume and temperature as follows: *The volume of a gas is directly proportional to its absolute temperature when the pressure remains constant.* This relationship was discovered independently by two French physicists, Jacques Charles and Joseph L. Gay-Lussac, and is now called *Charles' law.* Mathematically it can be expressed as:

$$\text{volume} = \text{constant} \times \text{absolute temperature}$$
$$V = k \times T \qquad (20.3)$$

or, if we let V_1 be a volume of gas at the absolute temperature T_1 and then change the temperature of the gas to T_2, the new volume V_2 will be given by the relationship:

$$\frac{V_1}{V_2} = \frac{T_1}{T_2}$$

By a similar experiment we can find that *the pressure of a gas varies directly with the absolute temperature if the volume remains constant.* The mathematical equation representing this concept can be written:

$$P = kT \text{ or } \frac{P_1}{P_2} = \frac{T_1}{T_2} \qquad (20.4)$$

where P_1 and T_1 are the pressure and absolute temperature of the gas at one time, and P_2 and T_2 the pressure and temperature under the new conditions.

Let us now consider the change the volume of water undergoes as the temperature varies. Figure 20-8 shows the

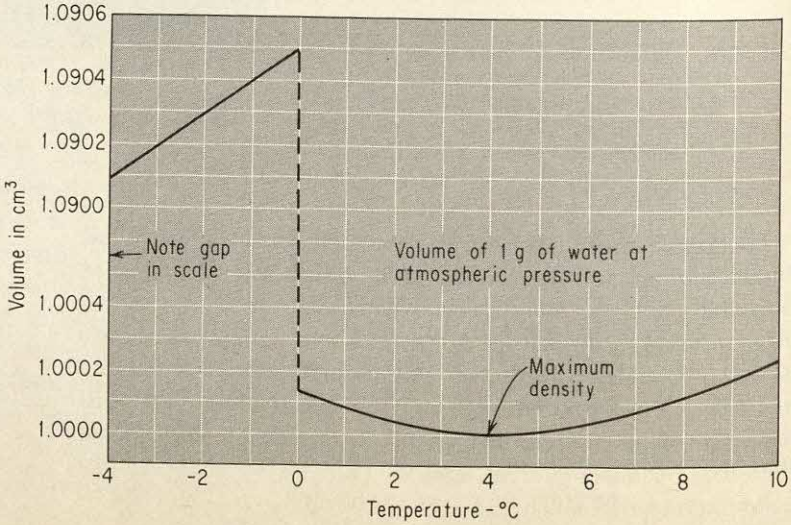


Fig. 20-8. Water's change of volume with a change in temperature. (From Linus Pauling, *General Chemistry.* San Francisco: W. H. Freeman and Co., 1953)

behavior of water at atmospheric pressure in the temperature range from −4°C to +10°C. Water is at a maximum density (or minimum volume) at 4°C and expands 0.013 per cent as it cools from 4°C to 0°C. It also expands as it warms above 4°C; it expands 1.19 per cent at 50°C, and 4.07 per cent at 100°C. Since warm water is less dense than cold water, it tends to rise or stay on the top. Cold water, on the other hand, is more dense and thus tends to settle to the bottom. This tendency for warm water to rise and cold water to settle is the basis for one of the great heat transferring methods—convection—to be described in Section 20-6. When the temperature of water drops below 4°C and the water starts to expand, it no longer increases in density and therefore no longer settles. The settling movement of the cold water ceases and the heat transferring method of convection fails to operate.

The property of water to expand when it cools as well as when it is warmed is unique. Water also has another equally important property—it expands as it solidifies. As water freezes, the molecules position themselves in a crystal structure that requires 9 per cent more space than when the water is a liquid. As an example, 11 cubic inches of liquid water at 0°C will form 12 cubic inches of ice. This expansion due to freezing is what causes frozen water pipes and car radiators to crack and break. Ice floats on water because of its expansion, and there are other important implications of this property of water that will be considered after we study some of the methods by which heat is transferred.

Other materials behave quite differently when heated, as is illustrated by the curves of alcohol and mercury on the graph in Fig. 20-9. As the temperature increases, their volumes increase uniformly. If the graph extended low enough to include the freezing point of alcohol or mercury, we would see that upon solidifying they shrink in volume even more.

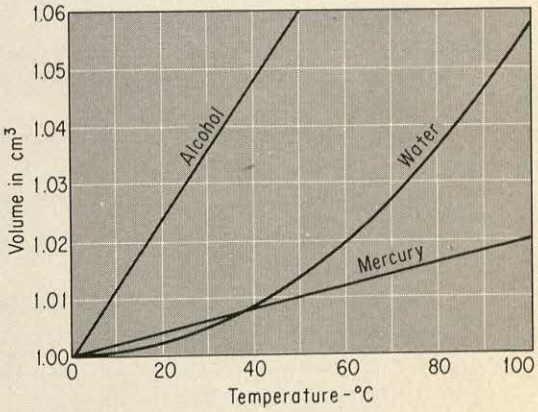


Fig. 20-9. Volume-temperature relationships for water, alcohol, and mercury.

20-6 Methods of Heat Transfer

There are only three ways by which heat is transferred: conduction, convection, and radiation.

Conduction is the flow of heat through matter unaccompanied by any motion of the matter other than atomic or molecular motion; that is, conduction is the transfer of heat by actual contact of molecule to molecule. When a material is heated the increased molecular motion is transmitted from one molecule to another by collision. There is no transfer of the material itself, and the energy of the increased motion of the molecules moves from the region of high temperature to the region of lower temperature. Substances vary greatly in their ability to conduct heat: in general, metals are good heat conductors while most other materials are poor conductors, or *insulators*, as we call them.

An experiment that was first performed by Count Rum-

ford gives us an indication of the ability of water to conduct heat and a clue to what conditions would be like if water behaved as other substances and contracted as it cools and solidifies. He put some ice in the bottom of a test tube, forced a small piece of wire gauze in on top of it to hold it in position, covered it with water, and applied heat to the water in the upper part of the tube (Fig. 20-10). Since the ice is kept down and since the heated water is less dense than the cooler water at the bottom, the water on top may be boiled for some time without any appreciable melting of the ice. This one simple experiment indicates that water is not a very good conductor of heat and that its density decreases as it is heated above 4°C. What if solid water—ice—always settled to the bottoms of lakes, rivers, and the ocean? The answers to such a question would lead us to picture nature in a vastly different way from what we do now. Water is indeed a key—and unique—substance in our physical world.

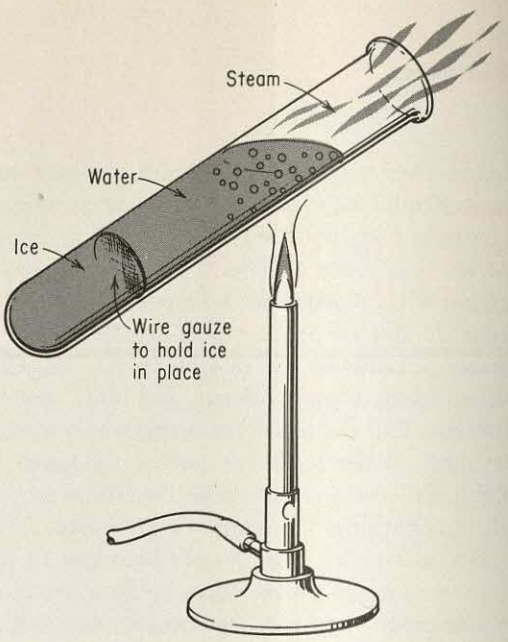


Fig. 20-10. Water is a poor conductor of heat.

The quantity of heat that is conducted by a material depends upon the following five factors:

1. The material itself. Silver, for example, conducts heat much more readily than does cork.

2. The area of material through which the heat passes. Twice as much heat can pass across two square feet of surface as across one square foot.

3. The temperature difference between two points in the material. More heat will be conducted between two points when the temperature difference between them is large than when the temperature difference is small.

4. Time. The longer the time in which heat can be transferred the greater the quantity of heat that is transmitted.

5. Thickness. The thicker the material the smaller the quantity of heat that can be transmitted in a given time.

From an analysis of these factors, we can conclude that the quantity of heat conducted through a substance is directly proportional to its cross-sectional area, the temperature difference between two points in it, and the time during which the heat passes. At the same time, the quantity of heat is inversely proportional to the distance through which it travels. We can state the above facts in a single formula:

$$H = \frac{KA(T_2 - T_1)t}{d} \qquad (20.5)$$

where K is the proportionality constant—which depends upon the nature of the material (Table 20-3)—A is the area, $(T_2 - T_1)$ the temperature difference, t the time, and d the thickness.

The apparent difference in the temperature of objects we touch outdoors in winter, such as iron, wood, or cloth, is often caused by nothing more than the difference in the rate at which these objects conduct heat away from our hands. Iron feels colder than wood because it conducts heat away from the skin almost 300 times faster than does wood. A material is often chosen to do a job because of its ability to

TABLE 20-3 **HEAT CONDUCTIVITY OF SOME COMMON SUBSTANCES**

	K^*	*Relative conductivity (silver = 1)*
Silver	1715	1.00
Copper	1615	0.94
Aluminum	890	0.52
Iron	285	0.17
Granite	20	0.012
Glass	3.35	0.002
Water	2.52	0.0015
Wood	0.90	0.0005
Sawdust	0.41	0.00024
Rock wool	0.26	0.00015

* K = Btu's transmitted over an area of one square foot per hour through a thickness of one inch for a temperature difference of 1°F.

conduct heat readily. Automobile radiators, cooking utensils, and steam and hot-water radiators are usually made of iron, aluminum, or copper because they are excellent conductors. Aluminum, you will note, is three times as good a conductor as iron, and copper is twice as good as aluminum.

In many cases, however, we want to prevent the transfer of heat, and then we must use a poor conductor as an insulator. Figure 20-11 compares the amount of heat lost in

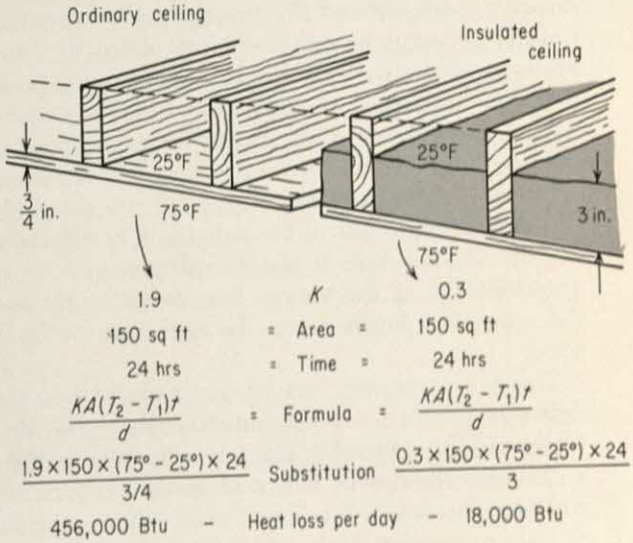


Fig. 20-11. Heat loss through a ceiling, uninsulated and insulated.

24 hours through an ordinary ceiling composed of plaster and lath having a "K factor" of 1.9 Btu/sq ft/hr/°F/in. with the amount of heat lost through a ceiling insulated with 3 inches of material having a K factor of 0.3. The room is presumed to be 10 by 15 feet in size and the air near the ceil-

ing 75°F and that in the attic 25°F. Considerable savings in fuel result from insulating homes, especially in northern states where temperatures may be low during several months of the year. In hot weather, conduction of heat from the outside into a house may be impeded by the same insulation.

Convection is the transportation of heat by moving matter, that is, by a flow of water or air currents. Hot-air furnaces and hot-water heating systems are examples of heating devices that utilize convection currents. When the material is moved by a blower or pump it is called *forced convection*, and when its flow is caused solely by the difference in its density it is called *free convection*.

To understand free convection, let us fill a U-tube with water (Fig. 20-12). Notice that the water level is the same in both arms of the tube. If we heat the right side of the tube, the water on this side expands, but the level in the colder left column remains the same, since it is balanced by the higher, but less dense, column on the right. If the stopcock is opened, water will flow from the warmer column to the cooler one, and this flow will decrease the pressure at the bottom of the warm column and increase the pressure at the bottom of the cold column. The water will then flow from the cold to the warm side around the bottom and a continuous circulation will be maintained as long as heat is applied to the warm side and removed from the cold side.*

Convection currents of both air and water are one of the most important methods of heat distribution over the earth's surface, as we shall see in Chapter 23.

Radiation is that process by which energy is transmitted through space without the presence of matter, that is, the transfer of energy by means of electromagnetic waves. These electromagnetic waves are light waves—radio, infrared, visible, and ultraviolet waves. If we touch a steam radiator, heat is transferred to our hands by conduction; when we hold our hand above the radiator, heat is transferred to it largely by air convection currents; if we place it at some distance from the side of the radiator, it is still warmed, and in this case the heat is transferred primarily by radiation. Practically all of the energy that comes to the earth from the sun or that finally leaves the earth does so through radiation.

The electromagnetic waves that transfer heat energy by radiation produce varying results at the receiving end. In plants and photographic film, these waves usually produce a chemical reaction or chemical energy. The radiation that strikes a photoelectric cell or a solar battery initiates a flow of electric current. (It would be helpful to review Section 12-4 at this point.) In a radiometer it generates mechanical energy and in other objects light energy that is visible to the eye. Radiation, therefore, is a process of energy transmis-

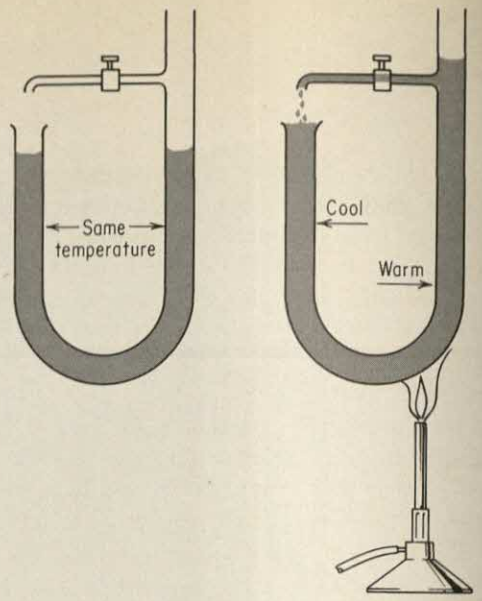


Fig. 20-12. Convection currents depend on the change of density with temperature. (From Sears and Zemansky, *College Physics.* Reading, Mass.: Addison-Wesley, 1953)

* Figure and discussion of convection adapted from Sears and Zemansky, *College Physics* (Reading, Mass.: Addison-Wesley, 1953).

sion that may or may not produce heat energy in the absorber. Radiant energy travels in straight lines, may be reflected by a mirror, can be bent or refracted by a lens or prism, and can produce interference effects like those of visible light. The amount of energy, and the relative amount of energy of short wave length, given off through the radiation process increases rapidly as the temperature increases (see Fig. 12-6).

To illustrate the relationship between radiant energy and temperature, let us see what happens to an ordinary kitchen stove. Even when the stove has cooled down to the temperature of our body, it will transmit radiant energy to us and will continue to do so as long as the molecules in it are vibrating, that is, until its temperature drops to absolute zero. The stove, of course, is also receiving radiant energy from all the other objects in the room. When the stove is at a higher temperature than its surroundings, it radiates more energy than it receives and its temperature drops; when it is cooled below room temperature, it will receive more radiant energy than it transmits and its temperature will rise.

To illustrate the relationship between wave length and temperature, suppose we watch a piece of iron being heated. It will first give off a dull red color that brightens to an orange or yellow as its temperature rises. This change in color is due to the fact that at higher temperatures an object gives off radiant energy of a shorter wave length. From this knowledge we can determine the temperature of a star by measuring the wave length that carries the greatest energy in the spectrum of that star. Contrary to what you might think, the bluer the light of the star the higher its temperature, and the redder its color the cooler it is. Figure 12-6 shows how radiated energy is distributed among the various wave lengths at different temperatures, and from it we can see that as the temperature becomes higher, the wave length that carries the maximum energy becomes shorter. The attempt to explain these curves theoretically led to the development of the quantum theory which was introduced in Chapter 12.

Objects with dull black surfaces are better absorbers and better radiators of heat energy than objects with shiny surfaces. A mirror reflects visible light rays, and its silvery surface also reflects radiant heat energy. In fact, the only effective insulation against heat transfer by radiation is a good shiny surface that will reflect the heat waves back toward their source.

Since there are only three methods of heat transfer—conduction, convection, and radiation—suppose we try to design a container that will eliminate heat transfer by any of these processes and enable us to keep food or any other material at its original temperature. We do not want heat either to escape or to enter. In order to eliminate heat transfer by conduction we could build a double-walled container to prevent direct contact between the outside wall and the

inner wall or the food. We could stop heat loss through convection currents by producing a vacuum between the two walls of the container. And, since a mirror reflects radiant heat, we can silver the walls of the container and eliminate heat transfer by radiation. Our device might look something like Fig. 20-13. We should not be surprised to learn that we are not the first to think of this "perfect" insulator. Such a vacuum bottle, which we know as a thermos bottle, was invented in 1885 by Sir James Dewar.

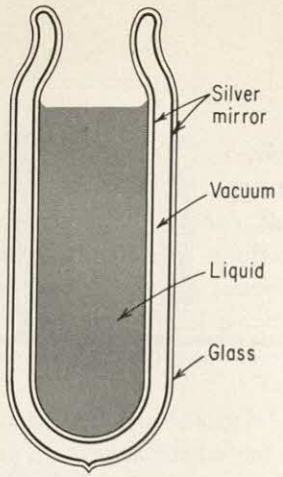


Fig. 20-13. A vacuum bottle or thermos.

20-7 Summary

The first big step toward understanding heat occurred when Galileo developed the thermoscope to measure temperature. Our present-day thermometers, however, are the result of the ideas of many scientists. A second forward step was taken when scientists recognized that temperature measures the *average* kinetic energy of the molecules of a substance while heat measures the *total* kinetic energy of a substance.

Some new terms that we encountered in this chapter are:

Calorie: the amount of heat required to raise the temperature of one gram of water 1°C.

Btu: the amount of heat required to raise the temperature of one pound of water 1°F.

Specific heat: the heat absorbed per unit mass per degree of rise in temperature.

Heat energy can be transformed into other forms of energy. For example, one Btu is equivalent to 778 ft-lb of mechanical energy.

Heating causes materials to expand; gases, for instance, expand uniformly by $\frac{1}{273}$ of their volume at 0°C per degree Celsius of rise in temperature. Water has the unique property of expanding both when it is heated above or cooled below 4°C and when it is solidified.

There are three methods by which heat can be transferred: conduction, actual molecule to molecule transfer of energy by molecular collisions; convection, actual movement of the warm molecules from one place to another by flow of a liquid or gas; and radiation, energy transfer by means of electromagnetic waves, light, heat, radio, and so on.

EXERCISES

A. KEY TERMS TO REMEMBER

absolute zero
British thermal unit
calorie
Charles' law
conduction
convection
fixed points
heat
radiation
specific heat
temperature

B. QUESTIONS ABOUT HEAT ENERGY

1. What disadvantages can you see in using water as a thermometric fluid?

2. In order to have a thermometer with the greatest spacing between degree marks, which liquid should be used, alcohol or mercury?

3. In what way does the high specific heat of water aid living organisms?

4. How does the high specific heat of water affect the summer and winter temperatures of the earth?

5. Suppose that water suddenly changed its behavior in the manner indicated below:

a. decreases its volume as it cools to the freezing point;
b. decreases its volume as it solidifies;
c. decreases its specific heat, as a liquid, to 0.1;
d. decreases its ability to dissolve substances by a marked amount.

For each of the cases above, considered individually, (1) what changes in nature would be noticeable immediately, and (2) what changes in nature would be noticeable after several years? Be sure to show the relationship between the predicted change in nature and the changed property of water.

6. For the best radiation of heat, what should be the color of a radiator?

C. PROBLEMS ABOUT HEAT ENERGY

1. How many Btu's are exchanged in the process of heating a gallon of water (8.3 lb) from 60°F to 180°F in 10 minutes?

2. What is the temperature that results from adding 400 cal of heat to 80 g of water, the temperature of which is 20°C?

3. What is the final temperature when 500 g of sand (specific heat = 0.23 cal/g-°C) at a temperature of 80°C are added to 100 g of water at 20°C? Ans. 52.1°C.

4. If 400 g of iron at a temperature of 120°C are added to 300 g of water at 20°C contained in an aluminum cup weighing 200 g, what is the final temperature?

5. A man weighing 220 lb decides to reduce his weight 20 lb by working. If the fuel value of his fatty tissue is 17,000 Btu/lb, how many foot-pounds of work must he do? Ans. 264,000,000.

6. If 1000 cm^3 of gas at 20°C are warmed to 80°C, with the pressure constant, what will be the new volume of the gas?

7. One hundred cm^3 of alcohol at 0°C will increase to what new volume at 40°C? What percentage increase is this? Ans. 104.8 cm^3; 4.8 per cent.

8. A room 20 by 20 ft in size has a ceiling 1 in. thick made of a material with a K factor of 1.9 Btu/sq ft/hr/°F/in. If the room temperature is 75°F and the outside temperature is 0°F, how much heat is lost in 24 hours? If the ceiling had been in-

sulated with 4 in. of sawdust, how much heat would have been saved in 24 hours?

9. What wave length is most closely associated with the emission of radiant energy from a body at a temperature of 3800°C? At 727°C?

10. What volume will 20 cm^3 of alcohol at 0°C have at 33°C?

11. If 20 calories of heat are added to 100 cm^3 of mercury at 0°C, what would be the final temperature? What assumptions must you make to solve this problem?

Twenty-One

The Kinetic Theory of Matter

In preceding chapters we have seen that matter is considered to be made up of molecules that are in rapid motion. The theory based on this view is called the *kinetic theory of matter*, and to see how this theory works let us examine the properties of matter more closely.

Is it possible that in iron the molecules move even though it is a solid? Is there a relationship between the pressure exerted by a gas and the motion of the molecules? Why doesn't matter just fly apart? Can the movement of molecules explain evaporation and melting, capillary action and pressure?

During the century following 1750, many men contributed isolated ideas that finally culminated in the formulation of the kinetic theory of matter. As early as 1738, Daniel Bernoulli had said that ". . . heat may be considered as an increasing internal motion of the particles." Count Rumford, too, as a result of his classic experiments, suggested that heat is motion. In 1780, Antoine Lavoisier and Pierre Simon Laplace added, "[Heat] is the sum of the products of the mass of each molecule by the square of its velocity." (We would now say that the total heat content of an object is equal to the total kinetic energy of the molecules that compose the object.) Then in 1808, John Dalton formulated his atomic theory of matter (p. 207).

21-1 Postulates of the Kinetic Theory

By 1811, Amedeo Avogadro was able to formulate his hypothesis: *Under similar conditions of temperature and*

pressure, equal volumes of gases contain the same number of molecules. Other workers, too, added their theories until now we can make the following six assumptions about matter:

1. All matter is composed of molecules.
2. All molecules of matter, above absolute zero, are in constant motion with an energy that is measured by the temperature of the body.
3. The molecules collide frequently with each other and, in the case of a gas, with the walls of the container.
4. The distances between the molecules of a gas is large compared with the size of the molecules themselves.
5. The molecules are perfectly elastic so that no energy is lost during a molecular collision.
6. The motion of the molecules is random, that is, they move haphazardly in all directions.

These assumptions constitute the kinetic-molecular theory of matter. Table 10.1 gives us some idea of the small size and high speeds of these molecules.

TABLE 10.1

MOLECULAR QUANTITIES AT 0°C AND STANDARD PRESSURE (76 cm OF MERCURY)

	Hydrogen	Oxygen	Carbon dioxide
Gram molecular weight	2 g	32 g	44 g
Number of molecules in one gram molecular weight	3.011×10^{23}	3.011×10^{23}	3.011×10^{23}
Number of particles per cm³	2.7×10^{19}	2.7×10^{19}	2.7×10^{19}
Diameter of each molecule	2.4×10^{-8} cm	3×10^{-8} cm	3.5×10^{-8} cm
Mass of each molecule	3.3×10^{-24} g	53×10^{-24} g	73×10^{-24} g
Mean free path †	1.8×10^{-5} cm	0.9×10^{-5} cm	0.6×10^{-5} cm
Number of collisions per sec. per particle	1.09×10^{10}	0.5×10^{10}	0.3×10^{10}
Average velocity	1.7×10^{5} cm/sec	0.5×10^{5} cm/sec	0.4×10^{5} cm/sec
Weight of 1 cm³	0.9×10^{-4} g	1.4×10^{-3} g	1.98×10^{-3} g
Volume of 1 gram	11.11 cm³	806 cm³	505 cm³

† Average distance traveled between collisions.

To answer a hypothesis such as the foregoing must be checked against all known facts. In other words, we must observe the behavior of matter and see how the kinetic theory fits that behavior. What happens, for instance, when a solid melts? When we apply heat to a solid, the kinetic energy and therefore the motion of the molecules in the solid increases. It is thought that most of the molecular motion in a solid is vibratory—that is, the molecules vibrate back and forth about a fixed position. At the melting point, this vibratory motion of the molecules becomes so large that their fixed positions are destroyed and their orderly arrangement is disrupted, with the result that the molecules are able to

break away from their fixed positions and the solid is said to melt. The movement of the molecules then changes from vibratory motion to translational motion, which is movement along straight lines. Since the kinetic theory satisfactorily explains the behaviour of the melting process, we can conclude that, in this one instance at least, the theory seems to be a valid one.

The molecules of a liquid move more and more rapidly as the temperature is raised, and when their energies become great enough, some of them break away from their neighbours and escape into the surrounding atmosphere. This process is, of course, the familiar one we call evaporation. The molecules that escape and leave their slower (cooler) comrades behind are the higher energy molecules, which explains why a liquid that evaporates rapidly is noticeably cooler than its surroundings. Liquids can flow because the rapidly moving molecules are able to move past each other easily. They can expand, therefore, to the full of gravity and can "flow" downhill or until the surface of the liquid becomes horizontal and the potential energy of the liquid becomes a minimum.

There are many other tests for the kinetic molecular theory and we shall consider them as we study more of the aspects of matter and its behaviour. We will find that the theory proves all the tests and satisfactorily explains all our observations and is therefore an indispensable tool to use as we begin to understand nature.

11-3 Intermolecular Forces

In both solids and liquids there are a great number of closely packed molecules (Fig. 11-1); in a cubic centimetre

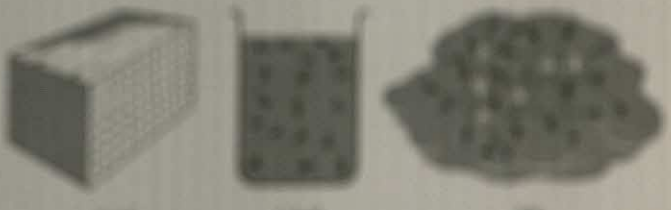


Fig. 11-1 The molecules (shown not accurately to size) in gas, liquid, and solid states.

Solid Liquid Gas

of average density there are approximately 10^{22} molecules, and the distance between them is in the neighborhood of $3 \times 3 \times 10^{-8}$ cm. It thus certainly does not seem unreasonable that there should be very strong forces of attraction on each other. In a solid, these forces hold the molecules in a relatively fixed position; in a liquid, they hold the molecules together in a definite mass and give the liquid a definite volume even though its shape may change. In addition, these forces cause a solid or a liquid to resist being pulled apart. In most solids the resistance can only be overcome by the application of

considerable force. The force required to break a rod is actually a measure of the force required to separate two layers of molecules within the rod. To pull apart the molecules of a steel rod an inch square, for instance, requires a force of 25 to 30 tons, and when we consider that each molecule weighs only 9.3×10^{-23} g and has an appreciable space between it and its neighbors, it is hard to believe that these tiny particles can exert such a large attractive force.

The attractive forces between molecules are of two types: those between like molecules (that is, molecules of the same substance) are known as *cohesive* forces. The force of 25 or 30 tons needed to separate the molecules in the steel rod is an example of cohesive force. The forces between unlike molecules are known as *adhesive* forces. A band-aid sticks to the skin because of an adhesive force between different kinds of molecules. Cohesive and adhesive forces both result from the same types of molecular forces we already considered in Chapter 14, that is, electrostatic forces, dipole interactions, gravitational forces, and other forces which we have yet to consider.

21-3 Surface Tension

The cohesive attraction between the molecules of a liquid produces what is known as *surface tension,* which is a contracting force that results in the tendency of all liquids to maintain as small a surface as possible. In many respects the surface of a liquid behaves like a stretched sheet of rubber. If a needle is carefully placed on a quiet surface of water, it will float, despite the fact that the density of steel is many times that of water. It floats because the water molecules are unable to "wet" the surface of the needle and the cohesive attraction of the water molecules is so great that they do not separate under the weight of the needle. If the surface of the water is broken, however, by the tip of the needle for instance, the surface tension will be overcome and the needle will sink.

To understand surface tension we must consider the attractive forces that act on the individual molecules in a liquid. In the interior of the liquid, the molecule is attracted in all directions by its neighboring molecules, and the attractive forces tend to balance one another (Fig. 21-2). At the surface, however, all of the "neighbor" molecules are alongside or below and the resulting attractive force on a surface molecule is downward.

The principle of minimum energy gives us another way to interpret surface tension. For instance, a ball rolls downhill, a piece of wood burns or, to generalize, things seek the lowest energy level. When a molecule is raised from inside a liquid to the surface, it works against the cohesive forces of the surface molecules, and potential energy will be stored in it in much the same way that a ball gains potential energy

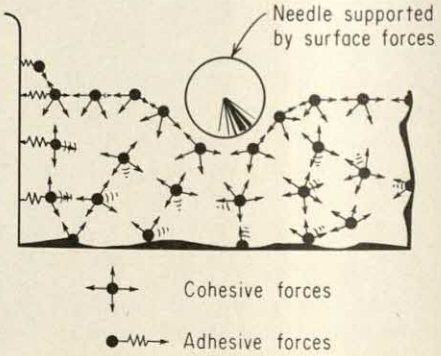


Fig. 21-2. Surface tension is the result of the cohesive forces between the molecules of a liquid.

when it is raised to the ceiling of a room. In order to reduce the potential energy to a minimum, the surface of a liquid tends to occupy as small an area as possible. If it were possible to have a raindrop suspended in mid-air, it would take the shape of a sphere because that geometrical shape has the smallest surface for a given volume. Falling raindrops are distorted because the surface tension is not great enough to overcome the air friction produced by their descent. Surface forces, then, tend to decrease the surface area and to form the shape of the free liquid. Liquids, in summary, are bodies of matter in which the molecules, though relatively free to move, produce a cohesive force strong enough to maintain a definite volume.

The relationship between cohesive and adhesive molecular forces is illustrated in the following experiment. A glass plate is suspended from one side of an equal-arm balance (Fig. 21-3) so that it hangs in a horizontal position. Enough

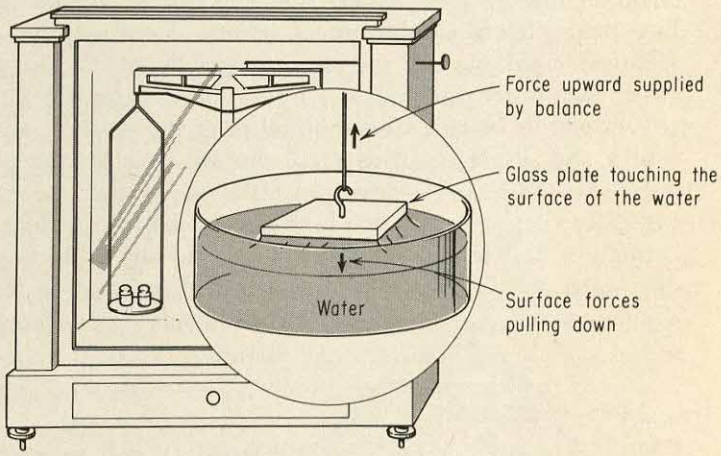


Fig. 21-3. An experiment for the measurement of the force of adhesion between glass and water molecules.

weights are added to the other side of the scale to balance the glass plate. A shallow dish of water is brought up under the glass plate until the plate is touching the water surface, and it is then lowered until the glass plate is very slightly below the level it was on before the dish of water was introduced. You might think the glass plate would easily pull away from the water; however, adhesive or "wetting" forces hold the glass to the water surface and our scales are no longer in equilibrium. We may add additional weights to the balance until the glass plate breaks free from the water surface. If we now examine the bottom surface of the glass plate, we will find that it is wet. The attraction between water and glass has caused a thin layer of water to be separated from the main body of water. This result indicates that the cohesive attraction of water molecules for other water molecules has been overcome but that at the same time the adhesive attraction of water for glass was strong enough to keep some of the water molecules clinging to the plate.

If we try this experiment with mercury in the dish instead of water, we find that it takes almost six times as much

force to separate the glass plate and the mercury. We also find that the glass plate is "dry"; that is, no mercury is attached to it after the separation. Thus, the adhesive attraction between mercury and glass is found to be smaller than the cohesive force between the mercury molecules themselves.

Water, of all the common liquids except mercury, has the greatest surface tension. This fact makes it of considerable interest in the field of biology. The rise of water within soils and plants, many of the effects of drugs on the protoplasm of different cells, and to some extent even the behavior of our blood are thought to be related to surface tension. And when a doctor prescribes ten drops of medicine, he knows exactly how much you are going to get because the surface tension of the medicine limits the size of each drop.

Sometimes it is desirable to lower the surface tension of a liquid. Soap, detergents, and many wetting agents accomplish this by their tendency to concentrate at the surface, acting to reduce the surface tension. Since soap has a relatively weak surface tension, it can enlarge its surface easier than water can, and, as a result, the soap or detergent will extend its surface around an oil or grease particle, wetting it and dispersing it as small droplets. Wetting agents are also used to help water penetrate fabrics that are being cleaned, and to aid dye solutions for fabrics. In lubricating oils, a detergent is used to assist the oil in "wetting" any water which may collect in it and disperse it as fine droplets rather than permitting it to accumulate as a sludge.

21-4 Capillarity

The adhesive and cohesive forces between molecules cause liquids to rise or fall in small tubes, and this action we call *capillarity*. If we stick a small glass tube into water, the water will rise in the tube. The adhesive forces between water and glass are great enough so that any water molecule that bounces up along the glass walls because of its ordinary random thermal motion will be strongly attracted by the wall and apt to adhere there. In Fig. 21-2, we saw a molecule adhering to the sides of the container in this fashion, with a water molecule just below it attracted to it by cohesion. As another molecule and still another move higher on the glass surface because of their random motions, adhesion causes them to stay there and cohesion pulls other water molecules along behind them. If we put a glass tube in water, the water will rise higher and higher on the side of the glass. Even though the molecules are very small and numerous—some 10^{23} of them per cubic centimeter—their motion is so rapid that the liquid seems to rise in the capillary tube immediately. Only under conditions where we could observe the molecules themselves could we see how capillary action really takes place, but nevertheless, the kinetic theory has

again given us a satisfactory explanation of an important observed fact in nature.

If we put mercury in a glass tube, any mercury molecules that rise above the surface will fall back, since the adhesive forces between mercury and glass are not as great as the cohesive forces of mercury. The mercury molecules, in other words, are more strongly attracted by the other mercury molecules than by the glass tube and this strong cohesive force causes the mercury surface to be depressed in the capillary tube and to round over in an attempt to form as small a surface as possible. Figure 21-4 illustrates the difference between water and mercury in surface tension and capillarity.

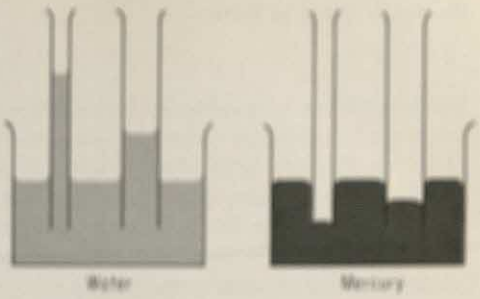


Fig. 21-4. Capillary action results in the rise of water and the depression of mercury in small tubes.

Many familiar examples of capillary action quickly come to mind—ink spreading in a blotter or in a cloth, water spreading in a towel, or coffee through a lump of sugar. More importantly, the distribution of water in soils depends largely on this same action. The spaces between the grains of soil form imperfect capillary tubes, which are large if the soil is coarse and small if the soil is fine. Through these spaces water moves from the deeper moist soil to the drier surface soil. When the surface of the soil is loosened by cultivation, the spaces may become too large for capillary action to carry water so readily to the surface. Cultivation therefore keeps the soil loose and reduces the loss of water through evaporation by reducing or preventing the rise of water by capillary action from the deeper soil layers. While most attention seems to be given to the rise of water in soils through capillary action, this is also very important in the movement of water downward and sideways through soils. Some dry soil in a clear glass container and a medicine dropper full of water can provide the basis of some interesting observations on the movement of soil moisture and on capillary action.

21-5 Osmosis

Suppose we put a film of cellophane over the large mouth of a funnel containing a sugar solution and then invert it and the cellophane under water (see Fig. 21-5). In a short time we will see the level of the liquid rise inside the funnel. The water is apparently able to move from the outside to the inside of the funnel through the cellophane.

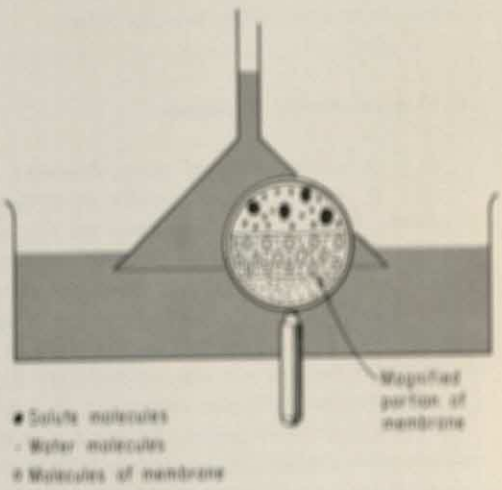


Fig. 21-5. A possible explanation of osmosis.

We can explain this movement—called *osmosis*—if we assume that the molecules of the cellophane film are spaced far enough apart to allow the water molecules, but not the larger sugar molecules, to pass through. Since the water molecules inside the funnel are interspersed with sugar and therefore have their freedom curtailed, more molecules of water will strike the bottom of the cellophane than the top. Thus, more molecules of water will enter the funnel than leave it. As the pressure inside the funnel increases, more

and more water molecules escape back through the film, and when as many water molecules pass one way as pass the other, equilibrium is reached. When this occurs the sugar-water solution in the funnel is higher and therefore exerts greater pressure on the membrane than the water outside the funnel. This pressure difference is called the *osmotic pressure* of the solution in the funnel.

Many plant and animal membranes act in the same manner as the cellophane in this experiment and allow water to pass rather readily in one direction. Dried fruits swell as they absorb water by osmosis during cooking. Fresh fruits, on the other hand, will shrivel when placed in a strong sugar solution. When a fried egg is salted, it is only a moment until each salt crystal is dissolved in the water that comes from the egg through osmosis. Too much rain will cause cherries on a tree to burst because water will pass through the skin of the cherry by this process. Meats will shrink when packed in salt, since their water passes out into the salt solution. Minerals in the soil that dissolve in water often pass through root membranes into plants. Human blood normally has an osmotic pressure of 100 lbs/in.^{2} but if we eat too much salt or sugar, which enters the blood, this pressure will rise and water will pass from the cells through the cell walls, causing us to become thirsty.

21-6 Behavior of Gases

It is in the behavior of gases, however, that the kinetic theory of matter is most vividly in evidence. Let us look again at some of the assumptions we are making about the molecules of a gas. These molecules are very small and very numerous. For instance, in a cubic centimeter of air at 0°C there are 27×10^{18} molecules, which is approximately the number of grains of sand in a cubic mile of sand.

These molecules are also exceedingly light in weight. A single nitrogen molecule, composed of two nitrogen atoms, weighs 4.8×10^{-24} grams. Although it is impossible to visualize such a small number, it may help to think of 4.8 grams divided into a million parts, with each of these tiny parts divided into a million pieces and in turn divided twice again into one-millionth of the remainder. Or, to put it another way, it takes 94 septillion—94 million million million million—nitrogen molecules to make a pound.

If there are 27×10^{18} molecules in one cubic centimeter of a gas, there must be $\sqrt[3]{27 \times 10^{18}}$ or 3×10^6 molecules along one edge. The centers of the molecules, then, must be separated by $\frac{1}{3 \times 10^6}$ or 33×10^{-8} cm. If the diameter of the nitrogen molecule is 3.15×10^{-8} cm, the molecules must be separated by approximately 10 diameters. If we compare this separation with the spacing between people in a classroom or on a dance floor, we can see that the mole-

cules are not crowded and that, in fact, a gas is mostly empty space.

We have already seen in Chapter 20 that gases expand when their temperature goes up. The second of our six assumptions about matter stated that as the temperature increases the kinetic energy of the molecules increases. As the molecules move more and more rapidly, the number and violence of the collisions between them increase. Those molecules along the outside edge of the gas collide more often and more violently with the walls of the container, and this increased bombardment of the container by the gas molecules results in an increased gas pressure or, if the gas is allowed to expand, an increased volume at the same pressure. A gas composed of such small, lightweight molecules can exert so much pressure because there are so many of them, moving at such high speeds, that an enormous number of collisions occur each second with the other molecules and the container's walls.

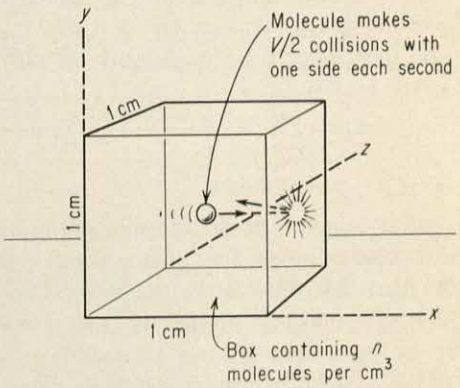


Fig. 21-6. The velocity of gas molecules in a box.

21-7 Velocity of Gas Molecules

From the pressure of a gas, we can determine the average speed of the molecules of gas at 0°C. Suppose we have a box 1 cm along each edge (Fig. 21-6) that is filled with nitrogen gas at 0°C and at one atmosphere of pressure. Since, according to the kinetic theory of matter, the gas molecules move in random directions, and the pressure exerted by the gas on one face of the box is the same as on each of the other faces, we should be safe in assuming that one-third of the molecules are moving back and forth in x direction, one-third in y direction, and one-third in z direction. Let us further assume that all the molecules are moving with the same speed, v cm/sec, and that they each have a mass, m grams. The pressure on the walls will be the same as if the molecules moved back and forth in straight lines between opposite walls without colliding with one another.

Each time a molecule strikes a wall of the container it loses a momentum mv in one direction and gains momentum mv in the opposite direction, and, therefore, at each molecular collision with a wall there is a change in momentum of $2mv$. Each molecule is going v cm/sec and the molecule will travel across the cube and back, a distance of 2 cm, in $2/v$ sec. In one second it will do this $v/2$ times. If there are n molecules in the cube and $\frac{1}{3}n$ move in x direction, the change in momentum at one wall is:

$$\frac{1}{3}n \times \frac{v}{2} \times 2mv = \frac{1}{3}nmv^2$$

This change of momentum gives rise to the pressure P on the side. Therefore,

$$P = \frac{1}{3}nmv^2$$

Since nm is the entire mass of the molecules in the cm^3 of gas, it is equal to the density d of the gas and we can write the equation:

$$P = \tfrac{1}{3}dv^2 \quad \text{or} \quad v = \sqrt{\frac{3P}{d}}$$

Remarkably enough, by the most elementary reasoning about the kinetic nature of matter we have computed the speeds of the molecules of a gas in the everyday terms of its pressure and density.

To calculate the speed of nitrogen molecules we have to know that nitrogen has a density of 1.25 kg/m^3 at 0°C when the pressure is one atmosphere or $1.033 \times 10^4 \times 9.8$ newtons/m^2. The velocity of a molecule of nitrogen, then, is:

$$v = \sqrt{\frac{3P}{d}} = \sqrt{\frac{3 \times 1.033 \times 10^4 \times 9.8}{1.25}}$$

$$= 493 \text{ m/sec}$$

This speed, which is about 18 miles per minute, is, of course, only an average speed, for some molecules move faster, some slower than this. The important thing, however, is the fact that by the simplest application of the kinetic-molecular theory we have been able to solve such a problem.

We have seen in Chapter 20 that the temperature is proportional to the average kinetic energy per molecule ($\frac{1}{2}mv^2$). At a given temperature, the average kinetic energy of a molecule of any gas is the same and therefore, as for any other gas, the average speeds of the different molecules must vary inversely as the square roots of their masses. Since a nitrogen molecule has 14 times the mass of a hydrogen molecule, the hydrogen molecule must move $\sqrt{14}$, or 3.74, times as fast as a nitrogen molecule, or at a speed of 1840 m/sec. Although these velocities of approximately 1 mi/sec may seem surprising, they have been carefully checked by direct experiments, and when such experiments agree with theory, we have added assurance that the kinetic theory of matter is correct.

21-8 Probable Distribution of Molecular Phenomena

We have determined the average velocity of the molecules of a gas, but we should not expect that all the molecules will have this average speed at any one time. Probably no two molecules move at exactly the same speed nor does the same molecule retain the same speed for more than about a billionth of a second between collisions. James Clerk Maxwell, who first examined the relationship between the number of molecules and their velocities, first considered the component velocity in a given direction and not the actual velocity with its three possible directions. Since as many

molecules rise as descend, the average velocity determined in such a fashion would be zero. By considering the speed in any one direction—such as up or down—and counting, for example, an upward speed of 1 mi/sec the same as a downward speed of 1 mi/sec, Maxwell showed that the distribution of a single component of velocity was similar to the probability curve shown in Fig. 21-7. Stated another way, if we called an upward velocity of 1 mi/sec positive and a downward velocity of 1 mi/sec negative, the average of the two would be zero. If we counted these same speeds without regard to the sign, we would have two particles moving at 1 mi/sec, and this type of consideration led to Fig. 21-7, which is the same curve we would obtain if we considered any chance happening.

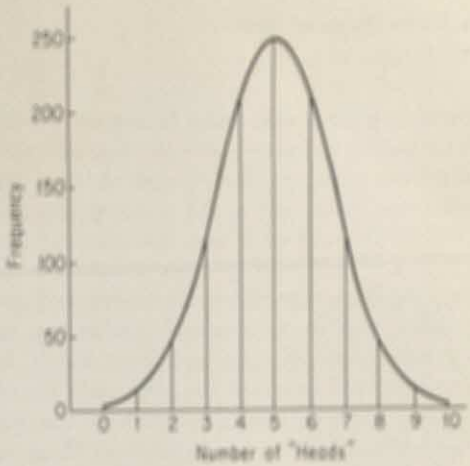


Fig. 21-7. A probability curve for a normal distribution, such as the frequency of "heads" in the toss of a coin.

If we toss two coins into the air, each once, we may get two heads or two tails, but it is more probable that we will get one head and one tail. If we toss the coins enough times, we will find that the number of heads will approximate the distribution shown in Table 21-2.

TABLE 21-2 EXPERIMENTS IN PROBABILITY

Tossing two coins

Number of heads	0	1	2
Number of occurrences per 100 tosses	25	50	25

Tossing ten coins

Number of heads	0	1	2	3	4	5	6	7	8	9	10
Number of occurrences per 1,024 tosses	1	10	45	120	210	252	210	120	45	10	1

By tossing more coins, we get a broader distribution that will eventually give the curve in Fig. 21-7, which is variously known as the *curve of normal distribution, error curve,* or *probability curve* and finds many applications in science, psychology, education, and social science. Such a curve is said to represent the relative frequency of occurrence of different combinations of a large number of similar, equal, independent factors when each of these factors is equally apt to occur. These conditions are often found in nature, and scientists utilize such a curve to predict the probability with which an event will occur.

Another type of distribution curve fits the behavior of gas molecules. When we consider the total velocity, which, of course, depends on all three of the velocity components (x

direction, y direction, and z direction), the distribution curve of molecular velocities is somewhat different than the previous one. Although the average of the x component of the molecular velocities or the y component or z component, taken by themselves, is zero, the chance that all three of the components will average zero at the same time is not great. Thus, we have to modify the curve in Fig. 21-7 to give a curve known as the Maxwell distribution law (Fig. 21-8). Some molecules go fast and some slowly, but as we have seen, the average speed depends upon the temperature.

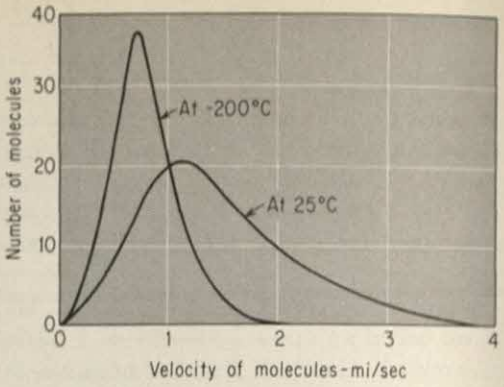


Fig. 21-8. Maxwell's distribution law applied to hydrogen gas.

Many problems can be solved from the distribution curves or laws that we have been studying. These problems include those dealing with the evaporation of vapor from a liquid, the dissociation of molecules into atoms, the "evaporation" of electrons from metals, and the "evaporation" of our atmosphere from the earth. Without the earth's gravity, the gas that comprises our atmosphere—the molecules of nitrogen, oxygen, carbon dioxide, and so on—would be scattered throughout space. By applying what we know about the distribution of molecular velocities, we can compute the energy that molecules must have in order to escape from the earth. Such reasoning indicates that hydrogen molecules in the atmosphere need 49 times as much energy as they have at present to escape from the earth's gravitational field. It also tells us that the moon cannot hold any atmosphere and therefore does not have twilight, wind or rain erosion, water power, or life of any kind upon its surface.

Although early scientists did not realize it, the laws they discovered were based upon the activity of the enormous number of molecules that make up the matter with which they were dealing. There is nothing that says molecules must be distributed according to the statistical laws we have been studying, for as we have seen, the position of the molecules and their velocities are a matter of chance. However, it is so very probable that the billions and billions of molecules will have speeds and behavior distributed according to these statistical laws, that science may deal with the motions of matter in bulk—in such fields as mechanics, heat and sound, and electricity, which deals with the average behavior of many electrons—and may assume that its predictions and conclusions, based on these statistical laws, are correct. The laws of probability are thus extremely important and enable us to apply the kinetic-molecular theory not only to individual events but also to mass behavior.

21-9 Brownian Movement

There are several direct ways of confirming the kinetic theory of matter. One of these was first discovered by Robert Brown, an English botanist who in 1827 observed that tiny microscopic particles suspended in a liquid (or in a gas) were in continuous motion. Fifty-two years later William

Ramsay, a Scottish scientist best known for his isolation of helium, explained the erratic motion of these small particles by suggesting that they were being irregularly bombarded by water (or gas) molecules. This random, quivering motion of these small particles is now called the *Brownian movement*.

In 1908, a French physicist, Jean Perrin, made a suspension of gamboge, a resin, precipitated in water. The resulting resin grains were spherical and of uniform size, and although these grains were too small to be seen by an ordinary microscope, he could follow them by watching the light that they scattered and thus could count their number and measure their size. After a long series of patient observations, he was able to determine the average change of position of these particles in one second, and from this he computed their average velocity. From the size and mass of these resin grains, he computed their average kinetic energy and deduced from this the mass of the molecule. Perrin received the Nobel Prize for this work in 1926.

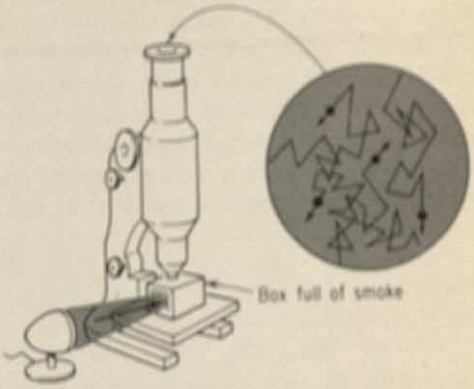


Fig. 21-9. Experimental arrangement for viewing the Brownian movement of smoke particles.

By using the smokebox in Fig. 21-9, into which a sample of air with smoke is drawn and then strongly illuminated, we can see the Brownian movements of smoke particles through the microscope. The observation of Brownian movement is one of the most direct and startling proofs of the motion of gas molecules.

21-10 Diffusion

Now let us turn to a second direct method of observing the kinetic nature of matter. We have all noticed how cooking odors soon spread through a house. The ability of a gas or liquid to spread throughout a space is known as *diffusion*. A gas can spread throughout an empty container and fill it almost instantly, and, even when the container is filled with some other gas such as air, the diffusion of the new or added gas still proceeds, although much slower, until the new gas completely and uniformly fills the container. At the speeds with which gas molecules move, it is no wonder that they soon mix with the molecules of any other gas to which they may be added. Even the motion of liquid molecules, although slower and much more affected by the gravitational fields of the surrounding molecules, is rapid enough and haphazard enough to allow diffusion to take place quite readily within the liquid. If a little coloring matter such as copper sulfate or food coloring is released at the bottom of a jar filled with water, the coloring matter will gradually penetrate to all parts of the water by diffusion.

To explain diffusion we must again apply the kinetic theory. Because the molecules of our liquid move back and forth, they gradually jostle some of the colored molecules to every part of the solution. In gases, with their higher molecular speeds, diffusion can take place more rapidly. Even in a

solid, with molecules that are more rigidly bound and in vibratory motion, diffusion can take place, although at a slower rate. That the molecules of a solid do move from one fixed position to another has been quite difficult to prove. Scientists are now using techniques involving radioactive materials and are finding that diffusion does take place. A block of iron is placed in an atomic pile for a period of time, at the end of which some of the atoms of iron will be radioactive and therefore detectable wherever they go. This block is then placed on top of a second iron block that has no radioactive atoms in it (see Fig. 21-10). After the blocks

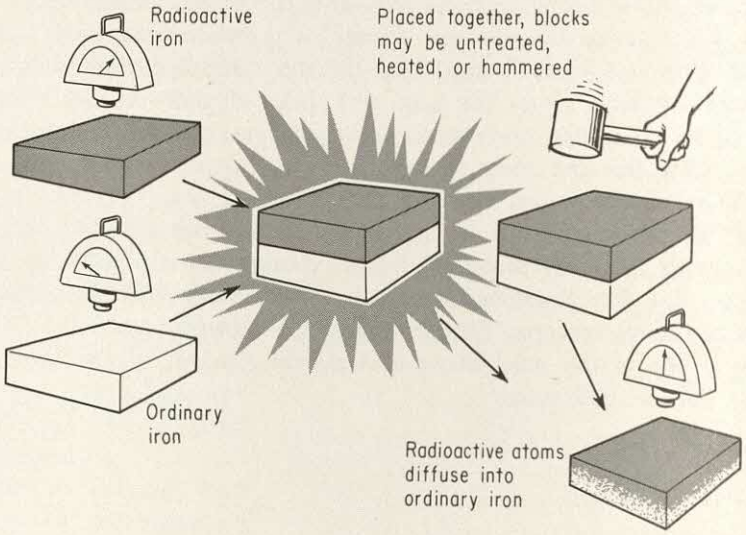


Fig. 21-10. An experiment in the diffusion of the molecules of a solid.

have been left in contact for a time, some of the radioactive atoms can be detected in the second iron block. Even if thin layers of the metal are shaved off, all of the radioactive iron will not be removed, for the radioactive molecules will have penetrated to some depth into the block composed of ordinary iron molecules. We must conclude then that the molecules of solid iron are in motion and that there has been an interchange or diffusion of iron molecules between the blocks.

If the two blocks had been heated while they were in contact, a still greater interchange of molecules would have taken place. This fact indicates that at a higher temperature there is a greater molecular activity, which confirms our definition of temperature in Chapter 20 based on the average kinetic energy of the molecules. We know that metal which has been hammered is warmed. The pounding gives the molecules of the metal additional kinetic energy, and so in a manner similar to that in the heated block of iron, some of the radioactive iron atoms diffuse into the solid block of ordinary iron. Even if we had placed the radioactive iron against the ordinary iron plate and left them untreated, there would still be a transfer of molecules from one to the other. This diffusion of molecules within a solid is further evidence

of the validity of the kinetic theory of matter. The radioactive tracer technique (considered by many scientists to be one of the most important tools in modern science) has fully justified our belief in the kinetic theory.

21-11 Colloids

In our common experience we encounter *solutions, colloids,* and *suspensions.* The properties of these three are distinctly different from one another; if we could magnify each of these to an extent necessary to see their particles, we would see why they differ. *Solutions* are dispersions of small molecules or ions in a continuous medium called the *solvent.* Salt dissolved in water, alcohol dissolved in benzene, and the carbon dioxide dissolved in soda pop are examples of solutions. A *suspension* contains particles easily identified with an ordinary microscope, and if we shake a mixture of sand and water, the sand persists as discrete particles too large to remain dispersed by the kinetic effects of the water molecules. Thus, most suspensions tend to settle out.

Between solutions and suspensions are *colloids* which contain dispersed particles in the size range of 2×10^{-7} cm to 5×10^{-5} cm. Solutions generally are limited to smaller particles and suspensions contain larger particles. There is no rigid boundary between these three systems, but whenever one or more dimensions of a particle (particles may be rodlike or thin sheets as well as spherical) lies in the colloidal range, certain unique properties are exhibited. Since the particles themselves may be solid, liquid, or gas, and they may be dispersed in any one of the three states, there are eight possible types of colloids as shown in Table 21-3. Notice that while gas dispersed in gas is listed in the table, its position there is qualified by indicating that a true solution will result rather than a colloid. The wide variety of examples in this table introduces us to the vast field of colloids.

An outstanding characteristic of colloids is their ability to scatter light. Although colloidal particles are too small to be seen in an ordinary microscope, these particles scatter light in all directions, and this scattered light may be examined

TABLE 21-3 TYPES OF COLLOIDS

Particles	*Dispersed in*	*Examples*
solid	solid	certain alloys
solid	liquid	gels
solid	gas	smoke, dust
liquid	solid	pearl, opal, mercury alloys
liquid	liquid	emulsions, milk, salad dressing
liquid	gas	aerosols, fog, fumes
gas	gas	(all gases are strictly soluble in each other)
gas	liquid	whipped cream, foams
gas	solid	gray hair, charcoal

at right angles to the incident light with a microscope. The scattered light appears much as the stars appear at night. In Fig. 21-9, colloidal smoke particles scatter light that is visible under the microscope. These tiny points of light appear to be in constant motion because the particles are constantly bombarded by molecules of the dispersing medium. This motion—the Brownian movement—is indisputable evidence of the existence of molecules. The scattering of the light by a colloid is known as the *Tyndall effect* and may be observed, for example, when sunlight passes through a dusty room or a searchlight points its slim finger of light into the night sky. Figure 21-11 pictures the path of a beam of light

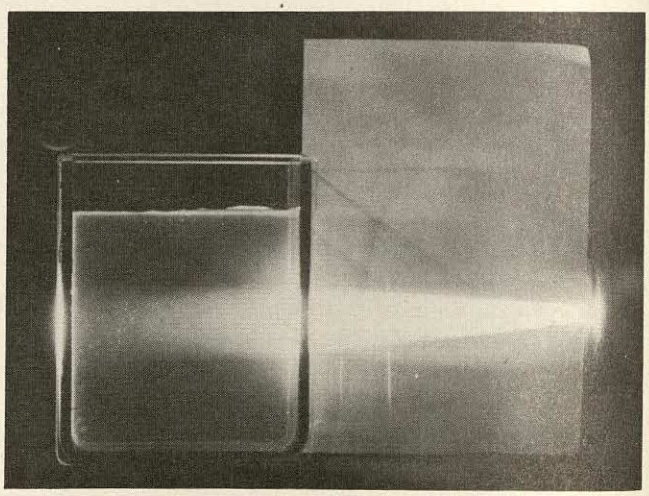


Fig. 21-11. A photograph of the Tyndall effect.

coming from a source on the right. Part of the light falls onto a white card, thus showing the path of the light beam. The tank contains a colloidal dispersion, and the path of the light through it is evidenced by the Tyndall effect. The size of the colloidal particles determines the wave length of the scattered light. The smaller the particles the shorter the scattered wave lengths. Consequently, white light may be sorted into its constituent colors, and certain colors may be scattered while others are transmitted.

The Tyndall effect often operates on a grand scale in nature. Since large particles scatter long wave lengths of light and small particles scatter shorter wave lengths, the sunlight that passes through our upper atmosphere is scattered by small particles and hence gives the sky a blue color, but the light from the evening sun comes through lower layers of the atmosphere where the dust particles are larger and thus the blue color gives way to the longer wave lengths of yellows, oranges, and reds. Because dust colloids are essential to colorful sunsets, sunsets differ considerably in their coloring in various parts of the world. The most magnificent sunsets are observed following major volcanic activity that pours out huge amounts of all sizes of dust particles, which scatter the sun's light and allow only the brilliant sunset colors to pass through.

A material in the colloidal state has a tremendous amount of surface area. To demonstrate this fact, let us examine a cubic centimeter of material. In this form the material has only six square centimeters of surface, but if it is broken up into colloidal particles that are 10^{-7} cm on a side, there will now be 10^{21} particles. Each of these small particles will have a surface of 6×10^{-14} cm^2, and the total surface area of the original cubic centimeter will now be 6×10^{-14} cm^2 $\times$ 10^{21}, or 6×10^{7} cm^2—sixty million square centimeters! The surface area of the cube has been increased ten million times, which gives our little cubic centimeter of material a surface area almost an acre in size. Figure 21-12 illustrates how

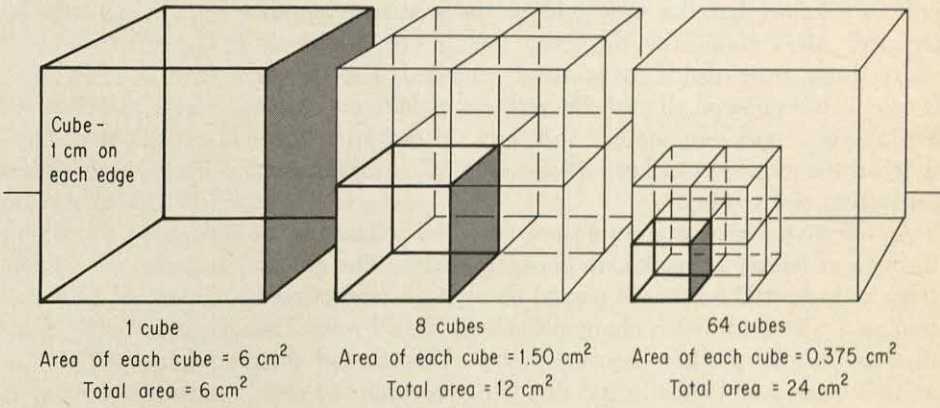


Fig. 21-12. The enlargement of surface area by divisions of volume.

surfaces are enlarged when the contents of a cube are divided into smaller and smaller units. Some of the most important properties of colloids stem from their enormous surface area.

When molecules are attracted by and stick to a surface they are said to be *adsorbed.* Many colloids tend to adsorb gases because of their large surface, others act as catalysts (Fig. 17-3), speeding up chemical reactions by bringing the reaction chemicals together on their surfaces. With the much greater area of the colloidal particles, there are more molecules exposed to possible chemical reactions and other types of activity. Much scientific research is devoted to the study of the surface phenomena that occur within the living cell. Protoplasm is colloidal and it seems certain that many of the chemical reactions and life functions taking place within the cell are due to the large surface areas of the colloidal substances.

Almost all colloidal particles carry an electrical charge, and since all surfaces tend to pick up a layer of charged particles, a colloid, which has a large amount of surface relative to the quantity of dispersed matter, tends to be highly charged. In order to examine the charge on colloidal particles, let us prepare a simple gold colloid. To a beaker full of boiling water we add a few cubic centimeters of for-

maldehyde and then a few drops of a dilute solution of gold chloride, $AuCl_3$. You will recall that gold is an inactive metal, easily reduced from its compounds to the free metal, but even a good reducing agent like formaldehyde fails to produce any visible precipitate of gold. Instead, a beautiful purple color—traditionally known as Cassius' purple—develops; this is the colloidal gold that a druggist displays in a large glass globe as a sign of his trade. If we place some of the colloidal gold suspension in the apparatus shown in Fig. 21-13 and pass an electric current through the colloid, after a few moments the purple color will disappear in the region near the negative pole. Since the gold particles are repelled by the negative pole and attracted to the positive pole, we know that the tiny gold particles are negatively charged. Most colloidally dispersed metals are negatively charged and their oxides are positively charged. The charge is usually the same on all particles within a colloid, and these like charges repel one another and thus aid the kinetic activity of the molecules in keeping the colloidal particles from separating and settling.

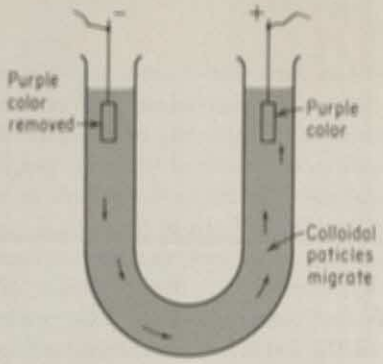


Fig. 21-13. An experiment to determine the electrical charge on colloidal gold.

An effective device to control smoke in cities is based upon the charge borne by colloidal smoke particles. The smoke from an industrial furnace is passed through a chamber containing a wire grid that is charged to about 50,000 volts. The dust and smoke particles become highly charged, and then as they pass near plates in the flue with the opposite electrical charge, they are quickly attracted to the plates and are thus removed from the gases that pass up the flue (Fig. 21-14). The precipiated particles often provide valuable by-products in addition to reducing the smoke nuisance of the industrial plant.

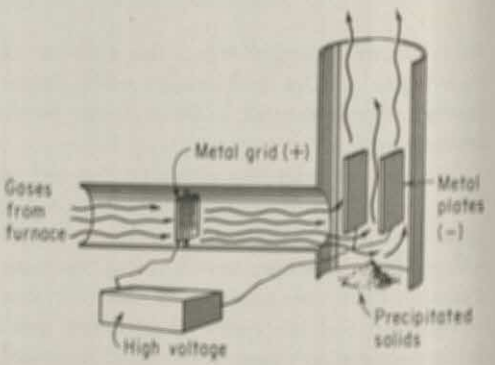


Fig. 21-14. The precipitation of colloidal particles by electrostatic means in a smoke eliminator.

Colloids, as we know, do not settle upon standing. The tiny, charged particles remain uniformly dispersed, even when attempts are made to filter them. Colloids, however, can be precipitated. In the gold colloid previously mentioned, the negatively charged particles mutually repel each other and hence remain scattered throughout the system, but if we neutralize the charge on these particles, the colloidal dispersion will collapse and the particles will settle. In addition to the direct neutralization of the charge by electrical means, the charges on colloid particles may be neutralized by adding salts or another colloid having particles of opposite charge. Once the charge on the particles has been neutralized, coagulation or precipitation occurs.

The coagulation of colloids plays an important role when river waters interact with the ocean at the river's mouth. The river, carrying suspended colloidal particles of soil, silica, organic matter, and other fine sediments, becomes mixed with the salt water, which neutralizes the adsorbed charges on these particles and causes them to coagulate. The resultant deposit of solids in the mouth of the river is just one of the causes which produces a delta. The deltas of such large rivers as the Colorado, the Mississippi, and the Nile

are examples of how the coagulation of colloids can help build up deposits over many square miles of area.

Two liquids that do not mix (said to be immiscible) may form a special type of colloid called an *emulsion*. If we add oil to water and then shake the resulting mixture, a milky disperson is produced, but upon standing, the oil and water separate into two immiscible or unmixed layers. However, if we add a small amount of soap to the oil and water mixture, the oil remains dispersed in the water. The soap is called an *emulsifying agent* because it has the ability to keep the oil and water dispersed. This emulsifying ability of soap enables us to wash oil and grease from our hands with soap and water. Milk is an emulsion in which the butterfat is dispersed in an aqueous system with the help of a protein emulsifying agent called *casein*. Olive oil is dispersed in vinegar and stabilized by egg yolk to make mayonnaise.

Gels are a unique kind of colloid. Certain substances such as gelatin, soap, albumin, and some clays, when dispersed in water, form gels upon standing. We are all acquainted with gelatin desserts, which are essentially a dilute dispersion (sometimes less than 2 per cent) of gelatin in water. Although the gel-forming agent is a colloidal particle, when this particle remains undisturbed for awhile it builds up a rigid structure with the water. The most popular theory of gel formation explains that the colloidal particles become associated with each other in long threadlike filaments that become tangled in one another. In this "brush-heap" structure the water is then held in the open spaces between the colloidal particles.

Certain rocks and minerals are gels. Agate, onyx, and flint are silicic acid gels that eventually became dehydrated. Seeds and some plant fibers may act as though they were dehydrated gels and, when they have the opportunity, absorb water, fill out the gel, and swell in size. Seeds such as beans absorb water to the extent that if a small hole is drilled in a log, filled with beans, and then stoppered, the beans will swell so much by taking up water from the surrounding wood that the log is likely to split.

Although we have been investigating extremely small particles—atoms and molecules and the strange world of the colloidal particle—we have found direct evidence of their existence in the Brownian movement, the Tyndall effect, and in diffusion. Since the kinetic-molecular theory greatly simplifies our understanding of many phenomena and since it seems to be a valid theory (we have yet to find a situation where it does not fit), we use it as the basis of our understanding of molecular motion.

In this remarkable sub-world we have been exploring, we have seen gas molecules darting in all directions at speeds varying up to 1 mi/sec. We have thought of them as tiny objects that suffer collisions, scatter throughout space, and move between other molecules, possibly of a different nature. So small and so numerous are these molecules that if

we were to make a hole in a perfectly evacuated light bulb so that one million molecules could enter each second, it would take 100 million years to fill the bulb with air. And yet these same small bundles of matter colliding with our hands like billions of tiny bullets exert the force that we call air pressure. The collisions of water molecules with the walls of their containers produce the water pressures or the steam pressures that turn our turbines and generate our electrical energy.

It is an amazing world, made up as it is of such small particles of matter. If we put a gram molecule of salt in the ocean and allow it to mix completely throughout the centuries, a glassful of sea water would still contain fifty of the original molecules. And yet, the evidence is such that much of science rests on this simple kinetic theory of matter. Now let us summarize what we know of this theory.

21-12 Review of the Kinetic-Molecular Theory

The six assumptions underlying the kinetic theory are:

1. *All matter is composed of molecules.* All matter—gases, liquids, and solids—is made up of these minute individual particles.
2. *All molecules of matter, above absolute zero, are in constant motion with an energy that is proportional to the temperature of the body.* We successfully used this hypothesis to explain the melting and evaporation of a material and found that diffusion and other properties of matter are consistent with this assumption.
3. *The molecules collide frequently with each other and with the walls of the container.* This assumption led us to the explanation for the pressure exerted by a gas and enabled us to measure the velocity of the molecules of a gas.
4. *The distance between the molecules of a gas is large compared with the size of the molecules themselves.* By knowing the number of molecules in a gas, we were able to determine that these molecules are approximately 10 diameters apart, and that, although in a cubic centimeter of gas there are 2.7×10^{19} molecules, matter is mostly empty space.
5. *The molecules collide elastically so that no energy is lost during a molecular collision.* We were unable to find any indication that the molecules of a gas lose energy during a collision, although we did watch small particles of matter bounce and move under the impact of molecular collisions or Brownian movement.
6. *The motion of the molecules is random; that is, they move in all directions.* By our statistical analysis, which was based on the fact that there are so many molecules,

we found that this assumption is a reasonable one, and from it we were able to derive the velocity of the gas molecules.

21-13 Summary

Much of our study in this chapter has been connected with the forces that hold matter together. The attraction between like molecules gives a force known as cohesion, while that between unlike molecules is called adhesion. These two forces result in capillarity, which is the degree to which a liquid will rise (or be depressed) in a small tube or space. Cohesion alone causes the surface of a "free" liquid to form a sphere. It also determines the size of drops and the shape of falling liquids. This pulling together of the surface of a liquid is called surface tension.

The random movement of molecules allows us to explain the Brownian motion. When we put a semipermeable membrane between a solution and water, we find that water molecules enter the solution until considerable pressure has been built up. This process is called osmosis and the pressure which results is known as osmotic pressure; both of these play a vital role in living things.

Molecular motion also allows us to explain the behavior of gases, how they diffuse, and how they exert pressure. Because there are so many molecules, we use the laws of probability to explain their behavior.

Solutions exist when the molecules or ions of one substance are dispersed throughout those of another material. Suspensions exist when one substance is not divided finely enough to remain dispersed but will gradually separate and settle. Colloids are the borderline case that exists between the suspension and the solution. The small particles reflect light giving the Tyndall effect, our sunset colors, and the haze that is caused by dust and smoke. Colloids are ordinarily highly charged and possess an enormous surface area.

EXERCISES

A. KEY TERMS TO REMEMBER

adhesion	colloid	osmosis
Brownian movement	diffusion	solution
capillary action	emulsion	surface tension
cohesion	gel	Tyndall effect

B. QUESTIONS ABOUT THE KINETIC THEORY OF MATTER

1. How are cohesion and adhesion alike? How are they different?

2. How is the resistance of molecular forces overcome? Give an example of how each kind of molecular force is overcome.

3. Why will a liter of water evaporate faster from a shallow pan than from one that is tall and narrow?

4. Give two ways in which the molecules of a gas may be made to come closer together. If these processes were continued, what would happen?

5. Do the molecules of a gas in a tank at high pressure necessarily have a high molecular speed? Explain.

6. Explain the action of cements, glues, and solders in terms of the molecular forces involved.

7. Why does a liquid rise higher in a narrow capillary tube than in a wide one?

8. In a gas composed of a mixture of light and heavy molecules, do the molecules all have the same kinetic energy? Explain.

9. If mercury is spilled on a surface of clean glass, it gathers into droplets, but water does not. Explain. What causes the water to form droplets when the glass surface is greasy?

10. What is the most recent example of capillary action, osmosis, and the Tyndall effect that you have seen?

11. Capillary action will cause water to rise 8 inches above the water surface in a particular type of glass tube. If we break the tube so that it is only 4 inches long, will water run out the top of the tube? Explain.

12. Recently there has been a controversy over our failure to use biodegradable detergents in industry and in our homes. (a) What is meant by biodegradable? (b) What is the present state of this controversy? (c) How much detergent is used in the U.S.?

13. Draw a sketch of an experiment that might directly measure the speed of molecules.

14. Actual gas molecules attract one another slightly. Would this tend to increase or decrease the actual pressure exerted by a gas compared to the pressure we might calculate if we were to neglect this factor?

C. PROBLEMS ABOUT THE KINETIC THEORY OF MATTER

1. A cube that is one foot on a side has what surface area in square inches? Ans. 864 sq in.

2. If the same cube is divided into cubes one inch on a side, what surface area will the material have? Ans. 10,368 sq in.

3. Toss two pennies 100 times and record the falls of the coins. Account for any discrepancy between your experiment and Table 21-2.

4. Predict the fall of three coins that are tossed 100 times and then test your prediction experimentally.

5. The bromine molecule has a mass 5.7 times greater than the nitrogen molecule. Determine the speed of the bromine molecule at 0°C and one atmosphere pressure.

6. In a gas made up of helium and oxygen (sometimes used in hospitals and by divers), (a) which molecules move faster? (b) what is the approximate ratio of their speeds?

Twenty-Two

Solids, Liquids, and Gases

The ancient peoples believed that matter was composed of four basic substances: earth, water, air, and fire. This concept, though naive and inadequate for our use now, represents an attempt to describe matter by its state. We may compare the concepts of earth, water, and air to the three states of matter that we now recognize: solid, liquid, and gas. Most of the things we encounter in our daily lives may be classed as a solid, a liquid, or a gas. Have you ever stopped to think how many gases or liquids you know? It is remarkable how few there are. As children we learned to distinguish the various states of matter, and as we study more about the properties of each of these three states, we will come to distinguish them even more clearly.

22-1 Solids

We find solids all about us—the ground on which we walk, the rocks on which we stub our toes, the minerals from which we obtain our metals, the wood from our forests, and innumerable animal and vegetable products. The very name "solid" is used as a synonym for unyielding and incompressible. Solids account for most of the earth's mass; 88 of the 102 elements of the periodic table (see p. 254) are solids at ordinary temperatures and pressures.

We have already seen that the molecules of a solid vibrate about a more or less fixed position. As a result of this limited molecular freedom, a solid does not flow but has a fixed shape and size. If a solid material is formed very slowly from the molten state, it may take a beautiful crystalline

form determined by the regular-patterned structure of the molecules. This regularity in turn forms crystals. We saw in Fig. 11-8 an example of this regular structure in a crystal of sodium chloride. All true solids are made up of one or more crystals. Whether the solid consists of a single crystal or, as is much more common, the solid consists of a large number of crystals, it is called a *crystalline solid.*

Some materials seem to have the properties of both solids and liquids. A piece of tar or pitch that is brittle and seems quite solid will, if placed in a cup, gradually assume the shape of the container, like a liquid. Sealing wax and silicone putty are other examples of materials that are not true solids for they have no definite permanent shape. They are sometimes called *amorphous* solids because their molecules are not arranged in definite patterns. Scientists usually consider them to be very viscous or slowly moving liquids rather than true solids. The properties of solids that we will now consider are those of the crystalline form of a material.

If a molten metal such as copper is cooled very slowly without jarring, it may form a few large crystals with most of its molecules in an orderly arrangement (Fig. 22-1). A crystal of copper is easily bent along one of its planes because the molecules will readily slide past one another. We saw how this can occur in the case of graphite and other crystals in Section 16-1. It is the regularity of the arrangement of the atoms in a crystal that gives to the crystal its characteristic properties.

Suppose we had a metal crystal in which the molecules could readily slide past one another, but we broke the structure into many smaller crystals with their molecular planes running in different directions. We could produce such a polycrystalline solid by adding impurities or by bending, jarring, hammering, or sudden chilling of the molten metal. Now when we try to bend the metal, any slippage of molecules cannot proceed very far before it is stopped by an adjacent crystal whose "slip planes" run in another direction. In just this way a metal may strengthen with use in a process called *work hardening.* A simple experiment will illustrate this: Heat a copper rod or piece of copper tubing and allow it to cool slowly. Notice that it will bend easily for it is made up of a few large crystals, but notice too that the more bending it undergoes the harder it becomes.

Crystalline materials have properties that are fixed; for example, they all have exact melting points. The regularity of the spacings of the molecules suggests that what happens to one molecule affects all of the molecules. Thus, crystalline solids are usually good conductors of heat, because when the molecules at one end of the crystal structure are heated and gain kinetic energy, their increased motions immediately affect the surrounding molecules. As the neighboring molecules gain kinetic energy, heat is transferred throughout the solid. There are many types of crystalline forms, and their

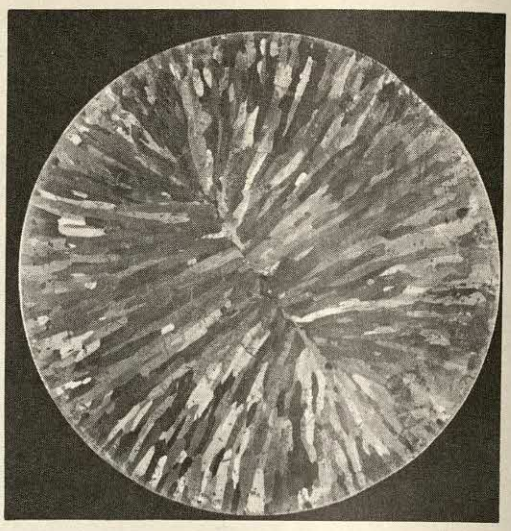


Fig. 22-1. This piece of copper was solidified and cooled slowly and then etched with an acid and polished. The photograph shows some of the many crystals that make up a metal. (Washington State University photo service)

properties largely depend upon the particular molecular pattern of their constituent molecules. Carbon, for instance, takes either of two crystalline forms, graphite or diamond (p. 310).

Solids are nearly incompressible at ordinary pressures, although some distortion does take place. They may be bent or stretched by appropriate forces, and if a solid springs back into its original shape after it is distorted, we say that it is *elastic*. According to the kinetic-molecular theory, the distorting force moves the molecules of the solid a short distance from their normal rest positions, but not far enough for them to overcome, even momentarily, the cohesive forces that bind them to their fellow molecules. When the distorting force is removed, the molecules return to their normal positions, and the solid springs back into its original form. If the distorting force is great enough, however, some of the molecules in the solid may be pulled or pushed out of their normal rest positions far enough to become entangled in the fields of force of other, "new," molecules of the solid. Then when the force is removed, the molecules find new rest positions within the fields of force of these new molecules and the solid becomes permanently deformed. When this occurs, the solid is said to have been subjected to forces beyond its *elastic limit*. To stretch mild steel permanently out of shape takes a force of approximately 35,000 lbs/in.^{2}, which

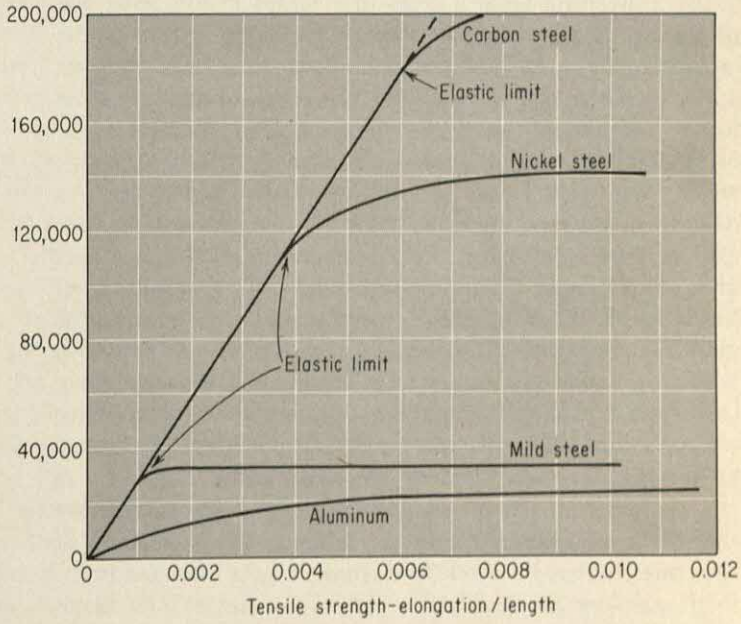


Fig. 22-2. A graph of stress vs increase in length for various materials.

is, then, the elastic limit of this material. The graph in Fig. 22-2 relates the change of length of various metals to the stretching force. Engineers need information of this kind in designing bridges and buildings, for they must know how

the materials they use will be able to withstand the forces that will be placed on them.

When a metal rod can be drawn into a very fine wire, we say that the metal is *ductile*. Gold, silver, and platinum are the most ductile of the metals. Platinum, for instance, may be drawn into a wire only $\frac{1}{20,000}$ of an inch in diameter. Ductility is, of course, the crucial property in the making of fine wires. The kinetic-molecular theory explains ductility in this way. The molecules of platinum, for example, can move past one another rather freely when a sufficient force is applied, but at the same time enough of a cohesive force remains to resist the ultimate separation of the molecules. Thus the material will flow but will not break.

A metal is *malleable* if it can be hammered or rolled into a thin sheet. In many respects malleability is similar to ductility. Whereas ductility is a measure of the molecules' ability to flow when subjected to a pull or tensile force, malleability is a measure of the molecules' similar ability with regard to compressional forces. Gold, silver, copper, and aluminum are among the most malleable of the solids. Gold is so difficult to break under compressional forces that it can be made into sheets that are nearly transparent. The chemical stability of gold coupled with its malleability make it valuable as a plating substance. Many solids, such as sulfur and salt, are so brittle that they cannot undergo much distortion of any kind without breaking.

The conversion from a solid to a liquid results from the increased kinetic energy of the molecules. As a few of the molecules increase their vibrational motion, due to heat added to the crystal perhaps, the interaction of the molecular forces and the regularity of the crystalline structure will cause all the rest of the molecules to be at, or nearly at, this same energy level. As a result, crystalline solids have a definite melting point, in contrast to other substances that do not have a definite melting temperature. When we heat glass, for example, it gradually softens and eventually flows like a liquid, but it is difficult to tell at what temperature the glass really melted. Such behavior occurs in substances in which the molecules do not form a regular pattern. Without the regularity of the crystalline structure, one small group of molecules may gain sufficient kinetic energy to behave as a liquid without affecting the neighboring molecules to any great extent. From our molecular theory we can predict that such amorphous materials, without any regularity of structure, cannot be good conductors of heat and electricity. With considerable justification, we can consider an amorphous solid as one that is intermediate between the solid and the liquid state. We can think of such materials as being liquids that are rigid, of high viscosity, and inelastic. The molecules in an amorphous solid are more independent of one another than they are in a crystalline solid, and, with a very small increase in this independence brought about by

the increased kinetic energy of higher temperatures, the amorphous material becomes a true liquid.

22-2 Liquids

Although liquids have a definite volume, they have no definite shape. The shape of a liquid's container determines the shape of the liquid, because, as we saw in Chapter 21, the molecules of a liquid are free to move past one another. The molecules in a liquid move more freely than those in a solid, and they usually are not as closely associated with one another as the molecules in a solid. Since most materials, on the average, are slightly less dense as liquids than as solids, greater spaces occur between the molecules of a liquid than of a solid. As a direct consequence of this, the molecules have a greater freedom of movement. Suppose we compare the molecules to persons in a crowd. If we happened to be standing in a crowd so tightly packed that there were no space around us, we could not move. If, on the other hand, there were a few openings in the crowd, we could move at least a little bit, for the person closest to the space could step forward, moving the hole backward, the next person could now step into this space, and so on. As the people in the crowd move forward, the spaces move backward, and one current theory holds that a liquid moves in much the same fashion.

Although we have seen that the molecules in a liquid are much freer to move than are those in a solid, the molecules cannot move indefinitely without collision. As a matter of fact, the molecules in a liquid usually have only a little more space to move in than do the molecules in a solid. The molecules, then, in a liquid simply "flow" outward and downward in any container until the liquid surface is horizontal and the potential energy is at a minimum. Because this property of flow is characteristic of both liquids and gases, they are called fluids. As the temperature of a liquid increases, it becomes less dense, and as its volume per unit weight increases, the amount of space between the molecules also becomes greater, allowing it to flow more freely. For this reason we use "heavy" thick oil in our cars in summer and "light" free-flowing oil in winter.

Liquids, as well as solids, are nearly incompressible because the space between molecules is so small. If this were not so, our hydraulic brakes, hoists, and presses could hardly function, for if the liquid absorbed the force itself, it could not transmit the force with nearly undiminished strength. Since liquids can be compressed only with great force, the density of a liquid remains almost the same regardless of its depth. Of some interest, however, is the fact that even though the compressibility of water is slight, if the water in the oceans were completely incompressible, the surface of

the ocean would be nearly 93 feet higher than it is at present. Water covers so large a portion of the earth's surface that the cumulative effect of its small compressibility is quite large.

Water is by far the most common naturally occurring liquid, although petroleum, molten lava, and some liquids produced by plants and animals do occur in small amounts in nature. You may be surprised to learn that only four elements of all those in the periodic table are liquids at normal pressures and temperatures. Even though water covers 70.8 per cent of the earth's surface and accounts for two-thirds of man's weight, it constitutes only 0.024 per cent of the mass of the earth. Water is so important to man that a loss of 20 per cent of the water in the human body will cause a tortured, painful death. Plants and animals alike depend for their very lives upon the 16,000,000 tons of water that fall (on the average) as rain on the earth each second.

22-3 Gases

Gases are familiar substances to all of us. We live at the bottom of an "ocean" of air, which is a mixture of colorless and odorless gases. We can "feel" the gas by waving our hand through the air or by putting it (carefully) out the window of a moving car. As "soft" as the air may feel to us under everyday circumstances, the impact of the air against the body of a jet pilot, unfortunate enough to have to bail out of his plane, is a serious hazard to life. Nevertheless, we use this same air which is so "hard" at jet speeds as a cushion against unwanted shocks when we encase it in various pneumatic devices.

We have already seen the kinetic-molecular basis for the behavior of gases. We know that the gas molecules move at rates up to one mile per second, make over a billion collisions a second, and have comparative freedom of motion. And we have seen how the volume of a gas varies with the pressure and how the pressure of a gas varies directly with the absolute temperature. But, as simple as the properties of gases seem and as common as our experience is with them, there is much that is difficult to understand about gases. It is hard to realize, for instance, that a tank full of oxygen will still be full of oxygen after the tank has apparently been emptied. If, for example, we have a tank full of 2 cubic feet of oxygen at a pressure of 700 lbs/sq in. and open the tank so that the gas can escape (by diffusing through the atmosphere), we will still have 2 cubic feet of oxygen in the so-called empty tank. The new pressure of the oxygen, of course, will be only one atmosphere, or 14.7 lbs/sq in.

Another peculiar fact is that gases have no free surface (like liquids and solids) and expand until they fill any sized or shaped container. So characteristic is this ability to expand indefinitely that it is the defining property of a gas,

and it seems to indicate that the forces between the molecules, which are large in the case of solids and somewhat smaller in the case of liquids, are practically negligible in the case of gases. This is in fact true, for the gas molecules are so far apart that there is almost no attraction possible between them. We have already seen how gases behave under varying pressures and temperatures, and in the next section we shall see still more completely how solids, liquids, and gases are related to each other.

22-4 Relationship between the Solid and Liquid States

To illustrate one aspect of the relationship that exists between the solid state, the liquid state, and the gaseous state, let us add heat to a substance at a uniform rate and see what happens. Suppose we heat a beaker containing ice at $-10°C$ with a steady flame and see how the temperature changes with time. When we plot the temperature as ordinate and the time as abscissa, we get a graph such as the one in Fig. 22-3. At first the temperature rises quickly be-

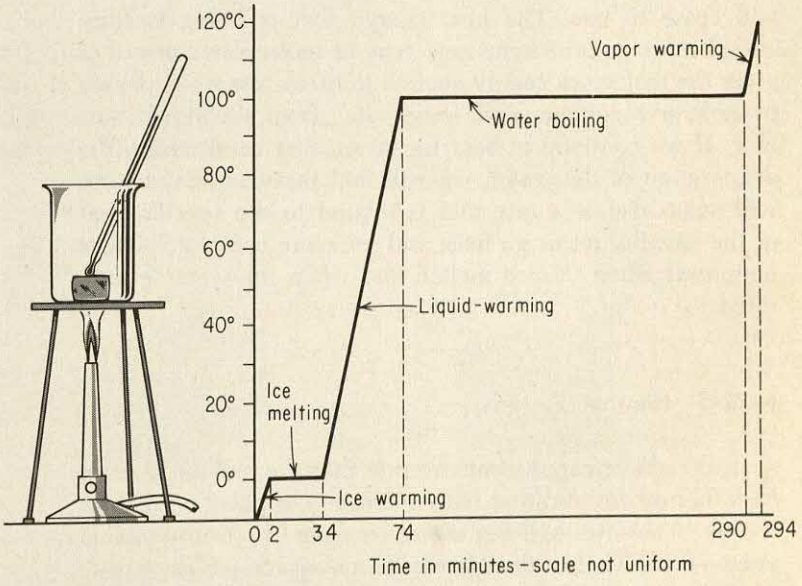


Fig. 22-3. An experiment in heating water.

cause the ice is being warmed and its molecules are moving more and more rapidly, but after 2 minutes liquid water appears and the temperature no longer rises. As time goes by, more and more ice melts as the molecules vibrate fast enough and far enough from their rest positions to interfere with their neighbors and, by collisions, to disrupt the orderly positions of the molecules of the solid. As the molecules break away from their fixed positions, the vibratory motion of the molecules is changed to a translational motion, and the solid melts. Thirty-two minutes after the ice reaches a

temperature of 0°C, the ice is completely melted, and the temperature again starts to rise. That the breaking apart of the orderly molecular arrangement of a solid takes energy is confirmed by the fact that the temperature remains constant while the ice is melting, even though heat is being added all the time.

The temperature of liquid water does not rise as rapidly as that of ice when heat is applied at the same rate. As the temperature rises, the translational motion of the molecules increases, and the increase in heat content of the water simply means that the kinetic energy of the molecules is increased.

We know from our study of Chapter 20 that a rise in the temperature of a solid depends upon the amount of material, the specific heat, and the rate at which heat is added. The same is true in the heating of a liquid, and if the rate of heat application is kept constant, as it is in this experiment, the rate at which the temperature rises will depend solely on the specific heat. If the specific heat is low, a large change of temperature will occur in a short time. If the specific heat is high, the change of temperature will be slow.

When the temperature of the water in our experiment rises to 100°C, the water will start to boil and the temperature will cease to rise. The heat energy that is being steadily added must go into some new type of molecular motion—it gives the molecules energy enough to break away completely from their neighbors and "evaporate" from the liquid surface. If we continue to heat the steam that results from the evaporation of the water, we will find that the temperature will again rise at a rate that is related to the specific heat of the gas. But let us go back and see more completely what happened when the ice melted and when the water evaporated.

22-5 Heat of Fusion

In the above experiment, we saw that the melting process took 32 minutes and that with a constant application of heat it took 40 minutes to bring the water from 0°C to the boiling point of 100°C. In other words, it took $\frac{32}{40}$ or 0.8 as long to melt the ice as it did to raise the resulting water to the boiling point. We know from Chapter 20 that it takes 100 calories to raise the temperature of one gram of water 100 Celsius degrees. To melt ice, then, with no rise in temperature must take 0.8 as much energy, or 80 calories. Since our experiment was crude, let us refine our procedure and reexamine the heat that is required to melt one gram of ice.

Suppose we have a glass beaker (specific heat of glass = 0.14) that weighs 300 grams and that contains 300 grams of water at 44°C. Now let us place 30 grams of ice at 0°C in the water and allow it to melt. If the final temperature of the water and the beaker is 34°C, how much heat was re-

quired to melt the ice? When we have to solve a problem like this, we can make use of the principle of conservation of energy, which, you will recall, says that the amount of energy going into a process is equal to the amount of energy that is returned. In this case, the amount of heat lost must equal the amount of heat absorbed by the ice and the water that resulted from the melting ice. Both the beaker and the water lost heat. The beaker lost an amount equal to its weight times its change in temperature times its specific heat or $300 \times (44 - 34) \times 0.14 = 420$ calories. The water lost heat according to the same formula: $300 \times (44 - 34) \times 1 = 3000$ calories. Altogether, 3420 calories of heat were lost by the beaker and the water, and this heat melted the ice and brought the resulting water up to the final temperature of the mixture (34°C). After the ice had been melted, the 30 grams of water that resulted from the melting had to be raised from the temperature of the melting ice to the final temperature (34°C), a process that took the following amount of heat: weight times change in temperature times specific heat or $30 \times (34 - 0) \times 1 = 1020$ calories. Since this heat was absorbed, it left 3420 — 1020 or 2400 calories of heat that must have been used just to melt the 30 grams of ice, without changing the temperature. Thus, it took $\frac{2400}{30}$ or 80 calories to melt each gram of the ice. The amount of heat needed to change one gram of a substance from a solid to a liquid, with no change in temperature, is called the *heat of fusion* of the material. For water, the heat of fusion is 80 cal/g or 144 Btu/lb. Figure 22-4 illustrates the fact that

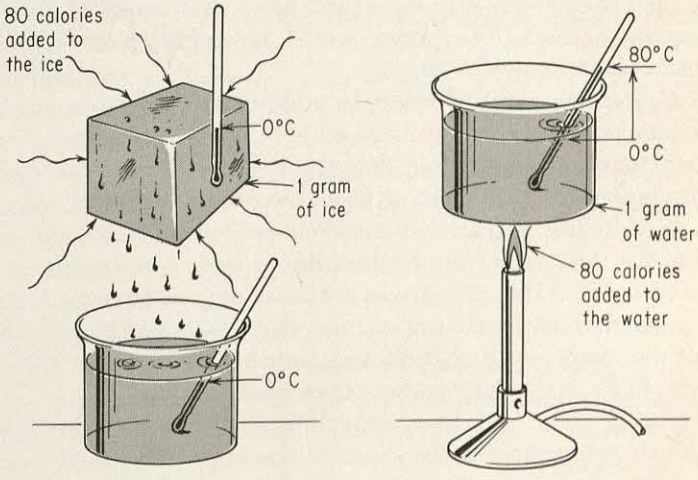


Fig. 22-4. Eighty calories added to ice at 0°C and to liquid water at 0°C results in different final temperatures owing to the heat of fusion of water.

80 calories added to 1 gram of ice at 0°C produces a different final temperature than it does when it is added to liquid water at 0°C.

Each substance that passes from the solid to the liquid state without decomposing or changing chemically has its own heat of fusion. Table 22-1 lists the melting points and

heats of fusion of some common substances. We must remember, of course, that the law of conservation of energy holds in the freezing and melting process, too, and that if a substance absorbs heat during the melting process it must also give it off while solidifying or freezing.

TABLE 22-1 SOME HEAT CONSTANTS OF MATTER

Material	*Melting point* °C	*Heat of fusion* cal/g	*Boiling point* °C	*Heat of vaporization* cal/g	*Critical temperature* °C	*Critical pressure* atm
Aluminum	658	76.8	2450			
Ammonia	−75	100	−33	330.1	132	111
Copper	1083	42	2300			
Ethyl alcohol	−112	24.9	78	204	243	63
Iron	1535	5.5	3000			
Mercury	−39	2.82	357	70	1550+	200+
Oxygen	−218	3.3	−183	60	−118.8	49.7
Water	0	80	100	540	374	217

From Table 22-1, it is apparent that the heat of fusion of water is high in relation to that of other materials. One of the noticeable effects of this fact is the long time it takes for ice to melt even in warm weather. Water's high heat of fusion is important in restraining spring floods, because as the increasing heat of the spring season raises the earth's temperature, that temperature is moderated by the absorption of heat in the ice-melting process. The high heat of fusion also moderates the earth's cold temperature, because as the air temperature drops to the freezing point, the water in lakes and streams starts to freeze and gives off an equivalent amount of heat. The lowest winter temperatures are tempered by this evolved heat.

The high specific heat of water, in addition to modifying the air temperatures near any large body of water, prevents the temperature of water from fluctuating widely. There is such an enormous evolvement of heat when large bodies of water start to freeze that a tremendous resistance is produced to the lowering of the temperature of large bodies of water below 0°C. This fact makes the freezing point something of a lower terminal temperature; that is, water tends to limit the temperature of itself and, through heat transfer methods, to limit the temperature drop of its surroundings to its freezing point (0°C). So, regardless of the severity of atmospheric temperatures, the ocean temperature will be a moderating influence on the climates of the earth.

22-6 Heat of Vaporization

If we refer again to Fig. 22-3, we see that at the boiling point of a material there is an absorption of heat energy just as there is when the material melts. It has been found that

as water condenses it gives off 540 cal/g, and when we evaporate water this same amount of heat is absorbed in the process. The amount of heat necessary to evaporate one gram of a material, with no change of temperature, is called the *heat of vaporization*. Table 22-1 shows that water also has a very high heat of vaporization in comparison with that of other materials. This property of water is even more important to life than the heat of fusion.

Wherever water is in contact with air, evaporation must take place until equilibrium has been established and the air is saturated with water. Unlike the case of freezing, however, the process of evaporation goes on continuously throughout all temperature ranges. The reason for this continual evaporation may not be clear until we think of the explanation of evaporation given by the kinetic-molecular theory. Within any liquid there are molecules that move more rapidly than the rest. Figure 21-8 indicated the wide variations that exist in the motions of gas molecules, and there is a similar distribution of molecular velocities within a liquid. By a series of "fortunate" collisions, a molecule may obtain a very high velocity. If it bounces toward the surface of the liquid with a great enough velocity, it will break through the surface and escape—or evaporate. Figure 22-5 shows several of these "fortunate" collisions that result in an escape and one that results in a hard collision with the wall of the container. Of course, there are molecules that return to the liquid as well as escape. If we cover the vessel, we will find that eventually there are as many molecules returning each second, on the average, as escaping, and when this occurs, the liquid and its vapor are in equilibrium.

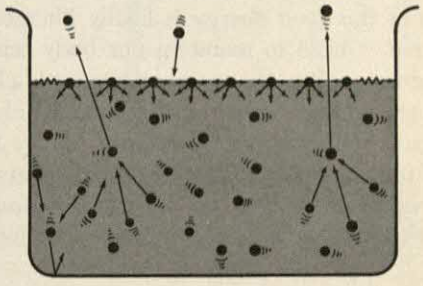
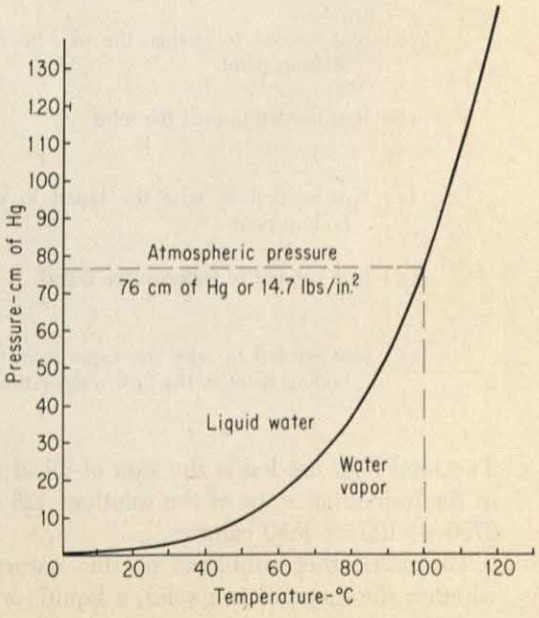


Fig. 22-5. Collisions between molecules may result in molecules escaping from the liquid to the vapor state.

The water molecules in the vapor state within the container, by their collisions with the walls of the container, exert a pressure that depends upon the number of water molecules and upon their velocities. As the temperature of the liquid becomes greater, the number of molecules escaping from the liquid (and their velocities) increase, and as a result, the pressure exerted by the vapor increases. This pressure is called the *vapor pressure* of a substance and it is dependent upon the material and its temperature. Figure 22-6 is a curve showing the vapor pressure of water at different temperatures. We can now define the boiling point of a liquid as the temperature at which the vapor pressure equals the atmospheric pressure.

Fig. 22-6. Vapor pressure curve for water, which shows the temperature at which water boils at various pressures.



The process of evaporation is an individual event, something that happens molecule by molecule. Because of this and because of the differences that exist between molecular velocities, evaporation from a liquid takes place as long as there is molecular motion. The evaporation of water, which entails the absorption of large quantities of heat, and the condensation of water vapor, which liberates the same amount of heat, are very important processes in the maintenance of uniform temperatures throughout the world. Water, because of the evaporation process, tends to resist

a sharp rise in the temperature of the earth when there is a sudden heat input such as occurs at sunrise. The large heat of vaporization of water is one of the things that makes water the most important agent in the transfer and control of heat on earth.

The heat of vaporization is also important in regulating our body temperature, for the evaporation of moisture from our skin absorbs large quantities of heat. When the air temperature rises above our body temperature on a hot summer day, we must depend largely upon evaporation for cooling, and we may evaporate as much as 1000 cubic centimeters of water per hour. The daily diet of the average American is 3000 kilocalories or 3,000,000 small calories per day. If one-fourth of this food energy is finally liberated in the form of heat that is used to maintain our body temperature, during the summer when the air temperature is close to 100°F our body temperature can be maintained only by evaporating approximately 1.5 quarts of water daily. Plants operate in much the same way. They evaporate water from their leaf surfaces in a process called *transpiration*, which in part makes it possible for them to survive a hot summer day.

22-7 Summary of Heat Exchange Problems

Now that we have studied the heat of fusion in the melting process and the heat of vaporization in the evaporation process, we are in a position to determine the amount of heat needed to make any change in temperature. The following calculations, of course, are valid only if no chemical change occurs during the heating. To review what we know about the heats of fusion and vaporization, let us determine the amount of heat in calories needed to raise the temperature of 5 grams of water from −50°C to 150°C:

(a) heat needed to change the solid to the melting point	= mass × change in temp × sp ht of the solid = 5 g × [0 − (−50)] × 0.5 = 5 × 50 × 0.5 = 125 cal
(b) heat needed to melt the solid	= mass × ht of fusion = 5 g × 80 cal/g = 400 cal
(c) heat needed to raise the liquid to the boiling point	= mass × change in temp × sp ht of the liquid = 5 g × (100 − 0) × 1 = 5 × 100 × 1 = 500 cal
(d) heat needed to vaporize the liquid	= mass × ht of vaporization = 5 g × 540 cal/g = 2700 cal
(e) heat needed to raise the vapor from the boiling point to the final temperature	= mass × change in temp × sp ht of the vapor = 5 g × (150 − 100) × 0.5 = 5 × 50 × 0.5 = 125 cal

The total heat needed is the sum of all of the heats needed in the individual steps of the solution: 125 + 400 + 500 + 2700 + 125 = 3850 calories.

To solve other problems of this nature, we determine whether the material is a solid, a liquid, or a vapor at both

the start and the finish of the process and then carry through the requisite operations.

22-8 Effect of Pressure on the Boiling Point

The boiling point—the temperature at which a material will normally change from the liquid to the vapor state—depends upon pressure as we can see in Fig. 22-6. If we take the temperature of boiling water on different days, we will find that the boiling point is not constant but changes as the air pressure changes. During evaporation, the volume of a material increases, and since pressure tends to prevent this expansion, an increased pressure handicaps the evaporation and, in effect, raises the boiling point.

To illustrate the effect of pressure on the boiling point, let us carry out the following experiment. Fill a round-bottomed flask half full of water and insert a thermometer, a pressure gauge, and an outlet tube through the stopper as in Fig. 22-7. When the water is first heated, small bubbles of air, which had been dissolved, start to form, rise to the surface, and escape. With more heat, bubbles of steam form on the bottom but collapse as they rise into the cooler water above and are condensed. As these bubbles of steam are cooled, the water vapor they contain condenses to liquid water again and produces for a moment a partial vacuum. As the walls of the bubble rush together and collide, we can sometimes hear a rattling noise or "singing."

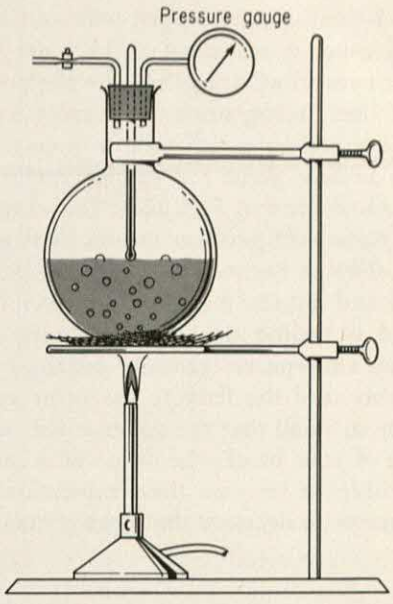


Fig. 22-7. An experimental arrangement for determining the boiling point of water at various pressures.

When the bubbles of steam are able to reach the surface of the liquid, the water is said to be boiling. As we watch what is happening inside the flask, we are unable to see the water vapor for it is transparent and colorless, but when it rushes out through the outlet tube and condenses, it forms the white cloud or mist of fine water droplets that we call steam. Although many people think that the white cloud of steam is water vapor, this simple observation shows that the vapor is actually colorless and transparent.

Suppose we partly close the outlet tube so that the pressure inside the flask will rise. As the pressure increases, the temperature of the boiling water increases, and this of course is what happens inside a pressure cooker. If we remove the burner and let the flask cool, the boiling will stop. But if we connect the outlet tube to an aspirator or vacuum pump and reduce the pressure in the flask, the boiling will start again, which confirms the fact that the boiling point is lowered when the pressure is reduced.

Another interesting way to demonstrate the reduction in boiling point with decreased pressure is as follows. Boil some water in a strong, spherical glass flask. When the water is boiling vigorously and after it has boiled long enough to replace the air in the flask with water vapor, stopper the flask and remove it from the source of heat. The boiling will stop almost immediately, but if we splash cold water over

the top of the flask, the boiling will start again (Fig. 22-8). The cold water causes some of the water vapor in the flask to condense and in doing so reduces the pressure in the flask. The fact that boiling starts again indicates that the boiling point is lowered by a reduction in pressure. If we carefully plot the boiling point for various pressures, we obtain a curve such as that in Fig. 22-6. The change in the boiling point of water with pressure creates some problems for cooks and housewives because it is the temperature that cooks the food and not the fact that the water in which the food is placed is boiling. If your flask were strong enough to withstand atmospheric pressure on the outside, you could conceivably cool the flask to the point where the pressure would be so small that the water could be made to boil by the heat of your hand. The value of a pressure cooker lies in its ability to increase the temperature of boiling water and therefore to decrease the cooking time for food.

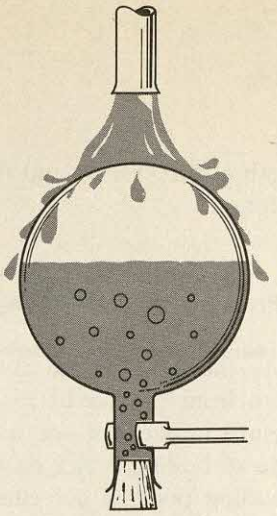


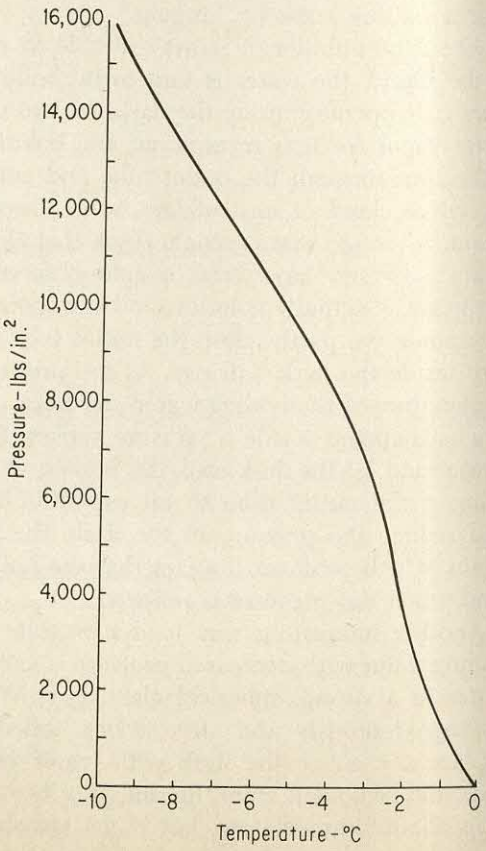
Fig. 22-8. Condensing the steam reduces the pressure on the surface of the water and allows the water to boil at a lower temperature.

22-9 Effect of Pressure on the Freezing Point

As the pressure on water is increased, the freezing point is decreased. Figure 22-9 is a graph of the temperature at which water solidifies plotted against pressure. Since water expands when it freezes, as we pointed out in Chapter 20 (Fig. 20-8), any increase in pressure tends to oppose this expansion, and the temperature has to be lowered further than usual in order that this change of state may take place. When we ice-skate, the increased pressure under the skate lowers the melting point of the ice below the temperature of the ice and we glide along on a thin, nearly frictionless layer of liquid water between the blade and the ice. We can see the streaks left by the refrozen water, and, when the temperature is just right, we may observe the liquid refreezing behind our skates as the pressure is removed. On exceptionally cold days, the pressure we exert may not be sufficient to lower the melting point to the temperature of the ice. When this happens the ice seems "sticky" because we no longer glide along on a water film. The flow characteristics of ice in large masses, as the glacier flow we studied in Chapter 4, may also be related to this pressure-melting point property of water. Water is an exceptional fluid in this respect, because most materials contract when they solidify and are thus aided in this process by an increase in pressure. For most materials, therefore, an increase in pressure increases the freezing point.

Since the boiling point of water decreases and the freezing point increases as the pressure is lowered, we may ask if it is possible for the boiling point and the freezing point to become identical at some pressure. It is in fact possible (Fig. 22-10) and occurs for water at 0.006 atmospheres (4.5 mm of Hg) and 0.0076°C. At this temperature and pressure,

Fig. 22-9. Variation of the freezing or melting point of water with pressure.



water can be seen boiling in the presence of ice. We call this the *triple point* of water.

An interesting demonstration will allow us to observe the behavior of water at the triple point. A good vacuum pump is used to pump out the bell jar arranged as in Fig. 22-11. Concentrated sulfuric acid is put in a dish in order to absorb some of the water vapor and to assist the vacuum pump in reducing the pressure. When the pump is started, the dissolved air in the water is released; as the pressure is reduced, water evaporates and cools the remaining liquid. As the pressure is reduced still further, with both the pump and the sulfuric acid carrying away the water vapor, the water should begin to boil. This very rapid evaporation uses up so much heat that the remaining water is cooled so rapidly that its surface eventually freezes, while at the same time the water under the ice is still boiling.

22-10 Sublimation

Only a few substances, of which water is one, have a triple-point pressure that is a small fraction of an atmosphere. The pressure at the triple point of carbon dioxide, for example, is approximately 5 atmospheres (at −56.6°C), and consequently this material cannot exist in the liquid form at normal atmospheric pressures. Solid carbon dioxide is called *dry ice* because, when it evaporates at atmospheric pressure, it changes directly into the vapor state and never becomes a liquid. This process of changing directly from a solid to a vapor is called *sublimation*, and is commonly experienced with iodine, camphor, and dry ice.

Just as there are temperatures that correspond to specific pressures in the boiling and freezing points of a material so there is a temperature-pressure relationship at which sublimation occurs. If heat is supplied to a solid, the solid will remain at the sublimation temperature until the heat of sublimation has been supplied and vaporization is complete. As we might expect, the heat of sublimation is approximately the sum of the heat of fusion and the heat of vaporization. Figure 22-10 illustrates all of these general relationships in the case of water.

Sublimation is more common in nature than might be supposed from the temperatures and pressures that are indicated for the triple point of water. Evaporation is, in effect, just the collective behavior of billions of molecules, and at nearly any temperature there may be one or more molecules that have experienced a series of advantageous collisions and are moving at a very rapid rate. A molecule may thus be moving as fast as a gas molecule and may break away from the surface and evaporate. It is in this molecule-by-molecule fashion that sublimation can take place from any solid. In some cases this sublimation is rapid enough to be quite noticeable. Clothes, for example, can "freeze dry" by

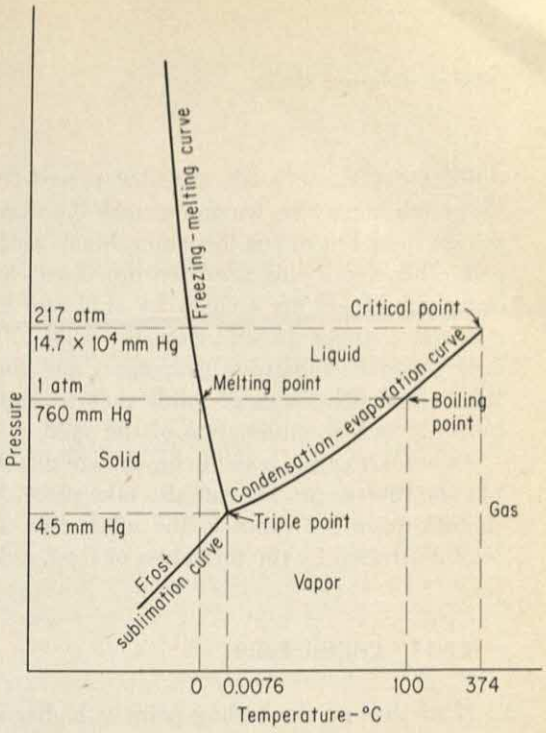


Fig. 22-10. Pressure-temperature diagram for water. Note that the scale is not uniform.

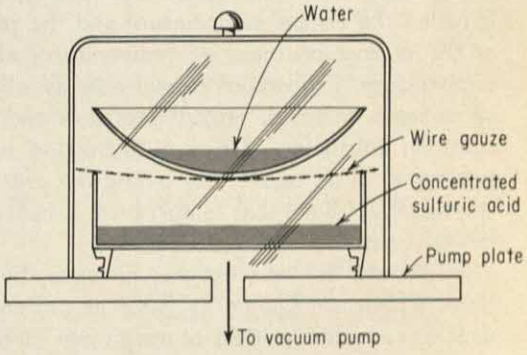


Fig. 22-11. Experimental arrangement for demonstrating water at the triple point.

sublimation. Some foods are now processed by letting the water sublime away, leaving a steak the same size and color as raw steak but minus the water. Many soup powders, eggs, and other foods are similarly processed. Sulfur and some other substances are refined by sublimation, for when the material changes directly from the solid to the vapor state, only the pure substance evaporates and the impurities are left behind. The odor of solids such as naphthalene (moth balls) is due to sublimation of the solid.

As we have seen, a solid can change directly into a vapor, but the reverse process can also take place. Matter may pass directly from the vapor to the solid state, a process that is well illustrated by the formation of frost and snow.

22-11 Critical Point

If we explore the boiling point at higher and higher pressures, we find an interesting thing. Above a certain temperature, it is impossible to find a liquid surface regardless of how much pressure we exert, although at a slightly lower temperature the substance apparently can be in a liquid state. It appears, then, that the condensation-evaporation curve is not indefinite in extent but has an upper limit, which is known as the *critical point*. The temperature at that point is called the *critical temperature* and the pressure is known as the *critical pressure*. At temperatures above the critical temperature, a substance cannot exist as a liquid. Table 22-1 gives some values of critical pressures and temperatures of common substances. Often a distinction is made between the vapor states above and below the critical temperature: the material above that temperature is called a gas and that below a vapor.

Substances such as oxygen, nitrogen, hydrogen, and helium, which are known as gases at ordinary temperatures, usually have low critical temperatures. Before one of these gases can be liquefied, it must be cooled to this temperature or below. For oxygen this temperature is −118.8°C; for helium −267.9°C.

We can observe what happens at the critical point with a simple piece of apparatus. Suppose we seal a heavy glass capillary tube full of solid carbon dioxide (dry ice). After it has been sealed and warmed to room temperature, the solid carbon dioxide changes to a liquid and is in equilibrium with its vapor inside the tube (Fig. 22-12). Since this experiment is hazardous, proper precautions must be taken, because the carbon dioxide at 20°C will exert a pressure of 830 lb/sq in. If we carefully raise the temperature, the liquid will start to boil and the meniscus (surface of the CO_2) will gradually become flat, then broaden and become hazy as the temperature approaches 31.1°C. As the critical point is passed, the surface of the liquid carbon dioxide vanishes. All the carbon dioxide is now a gas. As it cools again to

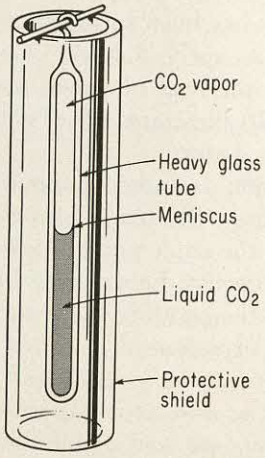


Fig. 22-12. Experimental arrangement for demonstrating the critical point of carbon dioxide.

the critical temperature, the meniscus of a liquid will again appear. As we might expect, the density of the liquid decreases as the temperature nears the critical point and the density of the vapor increases. At the critical point, their densities are the same (Fig. 22-13). As the temperature rises and approaches the critical temperature, the heat of vaporization decreases until at the critical temperature it is zero.

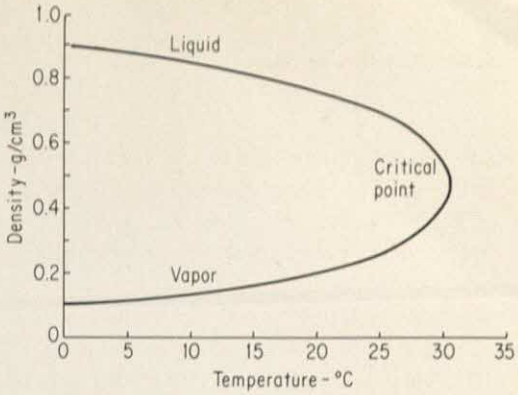


Fig. 22-13. The density of carbon dioxide near the critical point.

22-12 Pressure, Temperature, and Volume Relationships

Suppose we look at a material similar to that pictured in Fig. 22-10 but from a different angle to see what happens to the volume of the material as the pressure and temperature are changed. To do this, let us enclose a vapor in a cylinder that has a piston in one end with which we can change the pressure on the vapor. If we keep the temperature of the material inside the cylinder constant at whatever temperature we choose, as the pressure gradually increases, the volume of the vapor will gradually decrease, as in curve *AB* in Fig. 22-14. When the pressure reaches a value *E* corresponding to point *B*, we find that the volume decreases and that liquefaction occurs although there is no increase in pressure. The piston has arrived at point *B′* and further movement causes the vapor to condense into a liquid. As the piston is forced into the cylinder toward *C′*, the quantity of vapor decreases and the quantity of liquid increases. When point *C* on the graph is reached and the piston reaches *C′*, all of the vapor has been changed to a liquid, and an increase in pressure is required before any further decrease in volume will occur. It is soon apparent that it takes a great increase in pressure before there is much decrease in volume, and the curve *CD*, representing the compression of the liquid, becomes very steep. This type of process, where the temperature is kept constant, is called an *isothermal process*, and the line *ABCD* is called an *isothermal*. Figure 22-14 also shows this experiment as it would appear on the graph of Fig. 22-10. Since the temperature is constant, the line *ABCD* is vertical, and points *B* and *C* coincide on the condensation-evaporation line at the point where condensation takes place.*

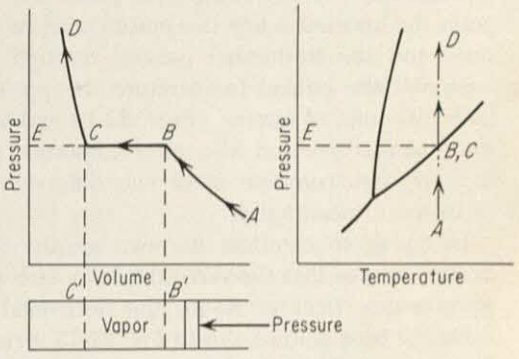


Fig. 22-14. Pressure - volume - temperature relationships during the isothermal compression of a vapor. (Adapted from Sears and Zemansky, *College Physics*. Reading, Mass.: Addison-Wesley, 1953)

As we carry this same sort of experiment out at higher and higher temperatures, we get a whole family of isothermals, as in Fig. 22-15. There are some differences in the curves, however, for we see that the horizontal portion of the curve becomes smaller and smaller until it eventually disappears and the isothermals form smooth curves. If we remember that the horizontal portion of the curve represents the conditions under which the vapor changes to a

* Discussion and figure adapted from Sears and Zemansky, *College Physics* (Reading, Mass.: Addison-Wesley, 1953).

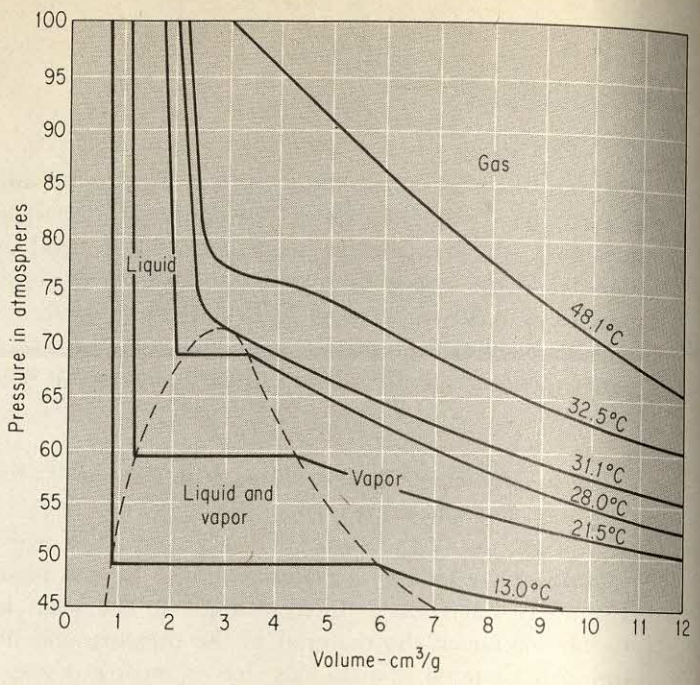


Fig. 22-15. Isothermal curves for carbon dioxide.

liquid, we know that, when the straight line disappears, liquefaction can no longer take place. The pressure at the point the horizontal line disappears must be the critical pressure, and the isothermal passing through this point must represent the critical temperature. It appears then that we have two sets of curves (Figs. 22-10 and 22-15) that have some similarities and also some important differences. Perhaps we can combine these two different graphs to make them more meaningful.

In trying to combine the two graphs (Figs. 22-10 and 22-15), we see that the vertical axis in each represents values of pressure. In Fig. 22-10, the horizontal axis represents values of temperature, but in Fig. 22-15 it represents volume. Since we really have three things to coordinate—pressure, volume, and temperature—we need a three-dimensional surface with each coordinate or dimension representing one of the variables. Figure 22-16 represents such a three-dimensional "graph." This particular graph is called a *P-V-T* surface and represents the thermodynamic surfaces of water.

From this combination of all the information that relates pressure, volume, and temperature to the states of matter, we can see that the triple point is really a line. The triple point has a fixed pressure and temperature but the volume

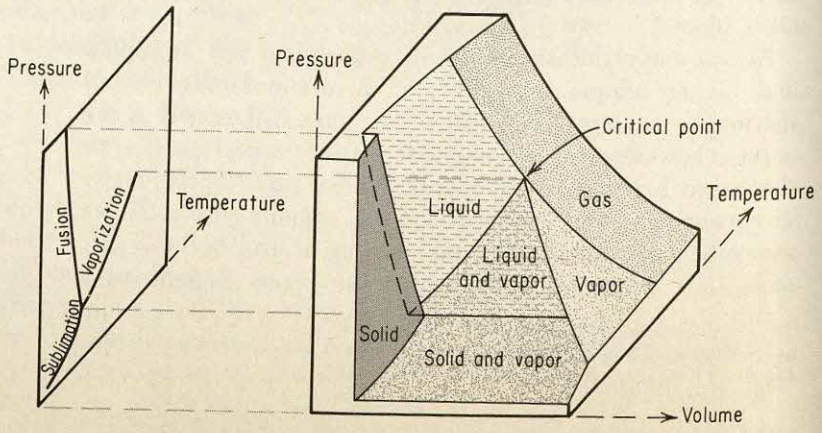


Fig. 22-16. Pressure - volume - temperature diagram for water.

will vary depending upon the relative amounts of material in the solid, liquid, and vapor states. It is interesting, too, to note that what is represented in Fig. 22-10 by the condensation-evaporation curve is represented in Fig. 22-15 by the area under the dotted line.

The amount of information contained in a small space by graphs of this type is indeed remarkable. Figure 22-16 shows us on one figure how the freezing point, boiling point, and sublimation point change with pressure, and we can deduce from this whether a substance will expand or contract in any of these processes. The diagram also gives information about the variation in volume of a gas as the pressure changes. It is no wonder that scientists use such graphs when they contain so much information in such convenient form. Although our graphs have dealt primarily with water and carbon dioxide, these substances were taken as typical examples of all substances that do not change chemically during the change from solid to liquid to gas. All such substances will have similar curves to describe their behavior.

All the information that is presented by these graphs was derived from experiment, but it is also of interest to the theoretical physical scientist to explain these curves as a result of the behavior of the atoms of the material involved. The explanations of these scientists have not always been simple, and the refinements of their theories are still being worked out. We have put an enormous amount of information in a small space on these graphs, but it is useful information only if you understand and can use the graphs. The questions at the end of this chapter will help you gain a better understanding of their use.

22-13 Summary

The understanding of solids, liquids, and gases is largely dependent upon our understanding of molecular behavior. When molecules are arranged in definite, relatively fixed positions, the substance is said to be a crystalline solid. Solids are nearly incompressible because the molecules resist closer packing. They have a definite volume and shape, and they resist bending and stretching because the cohesive forces between the molecules are so great.

The elastic limit is reached when some of the molecules of a solid have been pulled or pushed so far from their normal rest positions that they become entangled in the fields of force of their new neighbor molecules.

Ductility is the ease with which a substance can be drawn into a fine wire. Malleability is the ease with which a substance can be hammered or rolled into a thin sheet. An amorphous solid is a substance in which the molecules do not form a regular pattern; it may be thought of as a viscous liquid.

Since a liquid has molecules that are free to move past

one another but that still remain attracted by the fields of force of other molecules, it has a definite volume and can flow, but has no definite shape. There are relatively few naturally occurring liquids, but one, water, because of its thermal and chemical properties, is essential to life.

Since a gas has molecules that are comparatively free, it does not have a definite volume or shape, but fills whatever container it is in.

As a material is changed by the addition of heat energy from a solid to a liquid and then to a gas, its molecules gain increased motion and freedom. The amount of heat needed to change one gram of a substance from a solid to a liquid is called the heat of fusion. The amount of heat needed to change one gram of a substance from a liquid to a vapor is called the heat of vaporization. The amount of heat needed to change one gram of a substance from a solid to a vapor is called the heat of sublimation.

The melting point, boiling point, and sublimation point change with pressure, as represented in Fig. 22-16, which is a rather complete summary of a great deal of experimental information.

EXERCISES

A. KEY TERMS TO REMEMBER

amorphous	elasticity	malleability
critical point	heat of fusion	sublimation
crystalline	heat of vaporization	triple point
ductility	isothermal	

B. QUESTIONS ABOUT SOLIDS, LIQUIDS, AND GASES

1. Defend the statement: "glass is a liquid."

2. Distinguish between solids, liquids, and gases.

3. What becomes of the heat energy that is absorbed by ice during melting?

4. Farmers often place tubs of water in a fruit cellar in order to prevent fruit from freezing during cold weather. Explain why.

5. Identify the probable state in which we would find each of the following:

Water at:		Carbon dioxide at:	
Temperature	*Pressure*	*Temperature*	*Pressure*
a. 80°C	1 atm	a. 40°C	75 atm
b. 0°C	10 atm	b. 30°C	65 atm
c. 400°C	0.5 atm	c. 20°C	60 atm
d. 100°C	300 atm	d. 32°C	60 atm
e. 0°C	300 mm Hg	e. 25°C	90 atm

6. What important difference is indicated in the pressure-temperature curve for carbon dioxide (Fig. 22-14) and for water (Fig. 22-10)? Explain the meaning and importance of this difference.

7. What will happen to a rod of nickel steel (cross-sectional area ¼ sq in.) when a load of 10,000 lb is applied? 20,000 lb? 40,000 lb?

8. In what ways are the heat properties of water extraordinary?

9. Is there any way of getting from the liquid to the vapor state without crossing the condensation-evaporation line? If so, describe the process.

10. Where on Fig. 22-16 could we find the graph given in Fig. 22-9? in Fig. 20-8?

11. Suppose water suddenly changed its behavior in the manner indicated below: (a) decreases its heat of vaporization to 10 cal/g; (b) decreases its heat of fusion to 5 cal/g. For each of the cases above, considered individually, (1) what changes in nature would be noticeable immediately, and (2) what changes in nature would be noticeable after several years? Be sure to show the relationship between the predicted change in nature and the changed property of water.

12. In the wintertime we can sometimes see a river frozen over from bank to bank above a dam, but for a short distance below the dam open water occurs. Give two hypotheses to explain this phenomenon in terms of the heat processes we have been studying.

13. From what you know of the substances (a) iron, (b) alcohol, (c) CO_2, (d) iodine, draw pressure-temperature curves similar to Fig. 22-10, putting the appropriate letter at the approximate position of the substance on the curve at room pressure and temperature. Even though you may not have numerical values, you do have evidence from your experience to show which way the curves are likely to slope.

C. PROBLEMS ABOUT SOLIDS, LIQUIDS, AND GASES

1. The volume of carbon dioxide decreases during liquefaction. How much decrease in volume will occur for one gram of carbon dioxide during liquefaction at 13°C? at 21.5°C? Ans. 83 per cent, 78 per cent.

2. Compute the heat of fusion of ice from the following data:

	Trial 1	*Trial 2*
Mass of beaker	400 g	100 g
Specific heat of beaker	0.1 cal/g-°C	0.214 cal/g-°C
Mass of water in beaker	300 g	200 g
Mass of ice added to water	20 g	10 g
Initial temperature of water	25°C	20°C
Initial temperature of ice	0°C	0°C
Final temperature of water	19°C	15.4°C

3. How much heat is needed to raise the temperature of a mass M of water from temperature T_1 to T_2?

	M	T_1	T_2	
a.	5 g	25°C	65°C	Ans. 200 cal.
b.	20 g	−20°C	140°C	
c.	4 lb	0°F	190°F	

4. (The following problem is no less real because of its fictional aspects.) In the course of recent experiments on Mars, a hitherto unknown substance, which was subsequently named bisbium, was discovered and samples were brought back to the earth for study. To examine the thermal properties of bisbium, 500 g were placed in an insulated container of negligible specific heat and allowed to reach an equilibrium temperature of 20°C. At this temperature, bisbium was found to be in the solid state. Heat was then applied to it at a constant rate of 50 cal/sec. The temperature and volume were measured continuously for a time of 100 seconds, and at the end of this time bisbium was observed to be a vapor. The pressure was held constant and the graphs of Fig. 22-17 were obtained. Answer the following questions using the data taken from either or both of the graphs of Fig. 22-17:

a. What is the melting point of bisbium?
b. What is the boiling point of the material?
c. How much heat was supplied during the first 100 seconds of the experiment?
d. What is the specific heat of the material in the solid state? the liquid state? the vapor state?
e. What is the heat of fusion? the heat of vaporization?
f. Would the solid bisbium float on liquid bisbium at the melting point as ice does on water?
g. Would an increase in pressure raise or lower the melting point of bisbium? the boiling point?

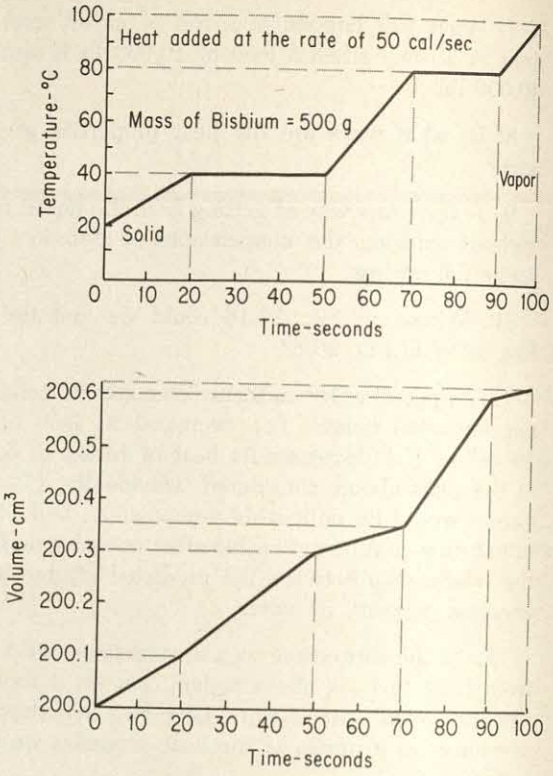


Fig. 22-17. Experimental data from the hypothetical element, bisbium.

Twenty-Three

Heat in Natural Processes: I

In these next two chapters, we will discuss the enormous importance of heat energy in our world and especially how this energy affects our own lives. We should remember, however, that heat is just one form of energy. If we were to discuss all the types of energy involved in natural processes, we would have to write a whole book, for the subject is almost unlimited.

Except for the energy that we obtain from nuclear sources in man-made devices such as nuclear reactors (see page 531), all of the energy available on the earth comes from the sun. Our fossil fuels are the remains of plants that were produced by the sun-powered process of photosynthesis or of animals that used plants as food. Water power is generated from water running downhill after it has been evaporated and lifted by the sun's energy. The vast convection currents in the ocean are the results of solar heat. The sun, to provide all this energy to the earth over a distance of 93 million miles, gives off energy at a stupendous rate—estimated by scientists to be 3.6×10^{23} Btu/sec. Since we are so far away, the earth receives only about one two-billionth of this energy, but even so, this amounts to a staggering 4,690,000 horsepower per square mile. Not all of this energy gets all the way to the surface of the earth, however, for 33 per cent is reflected by the clouds and another 9 per cent is scattered back into space by the atmosphere. The fraction that finally gets through to the surface is just enough to make life possible on the earth. But what is this life-giver like? Let us take a close look at the sun's structure and features, as diagrammed in Fig. 23-1.

23-1 Our Star—the Sun

The *corona* or outermost part of the sun appears as a pearly halo during a total eclipse. It may extend as high as 50 million miles, and its atomic particles move so rapidly that if temperature really had meaning in matter that diffuse, we would have to say its temperature is as high as 500,000°C. Scientists are still trying to find out exactly what the corona is. One of the current theories pictures it as clouds of hydrogen gas being gathered from outer space by the moving sun. This hydrogen is then later used in the nuclear reaction by which the sun maintains its high energy output. The sun, in other words, may be sweeping up its own fuel as it moves through the universe. Below the sun's corona is a region called the *chromosphere,* which is a "layer" that seems to be made of individual filaments up to 10,000 miles in length and to have a pinkish tinge when seen during a total eclipse. Scientists think that this region is also made of hydrogen and has a temperature of up to 20,000°C. The chromosphere can be thought of as the atmosphere of the sun.

The *photosphere* makes up the next layer and is what we see when we look at the sun. It has a temperature of approximately 6000°C and is thought to be made up of hydrogen atoms that have temporarily acquired an extra electron. The pressure at the top of this layer is estimated to be 0.15 lbs/sq in. or 1/100 of our atmospheric pressure. It is in this layer that the *sunspots* occur, and while these spots appear dark to the eye, they are only slightly less brilliant than the rest of the solar surface. The spots are strongly magnetic, some of them having fields 500 times as great as the earth's magnetic field. They may vary from 500 miles in diameter to an area of several million square miles, and they seem to occur in cycles of 11.4 years' duration. We do not have an adequate explanation for this sunspot cycle; it constitutes one of the many unsolved problems of science. Attempts have been made by many people to link the sunspot cycle to variations in the weather, rainfall, the stock market, and the birth rate, but so far no direct causal relationship has been shown to exist between them. There is, however, a direct and immediate deterioration of electrical communication on earth whenever a sunspot occurs, and this seems to be marked proof of the magnetic and electrical nature of sunspot activity.

Often associated with sunspots are large *prominences,* which seem to be clouds of glowing gas that may remain stationary or may rush upward as high as 250,000 miles at rates up to 100 mi/sec. The prominences may last from an hour to several days and may contain as much as one billion tons of gas. The sun itself has a mass of 2.1×10^{27} tons, which is 333,400 times greater than the earth, and a density of 1.42 g/cm^3. With a density as low as this we should not be surprised that the spectroscope shows it to be composed of 81.76 per cent hydrogen, 18.17 per cent helium, and only

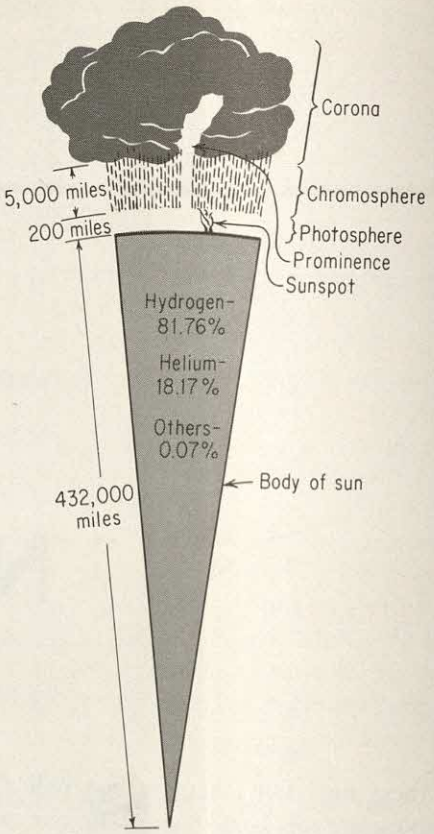


Fig. 23-1. A cross-section of the sun. (Adapted from Ayres and Scarlott, *Energy Sources—The Wealth of the World.* New York: McGraw-Hill Book Co., 1952)

0.07 per cent of other elements. We should not think of this hydrogen in the ordinary sense, however, for at the high temperatures existing in the sun it certainly has difficulty in maintaining a covalent bond with another hydrogen atom to form H_2, and in many instances it has been stripped of its electron. Each second in its nuclear reaction, the sun converts 560 million tons of hydrogen into helium, and in the process 4.2 million tons of hydrogen are transformed into energy. We shall consider in detail the mechanism by which the sun releases this energy in Chapters 25 and 26.

Regardless of how it is produced, the sun's energy supplies us with the heat and the energy for life on earth. It is the sun that gives the earth the energy for photosynthesis, rain, water power, and the unequal heating and cooling which causes the currents in the atmosphere and in the oceans. To see the full importance of the sun's energy for our lives, let us imagine that the sun were suddenly extinguished. Eight minutes later the earth would be plunged into darkness. The first cooling would be similar to that at night, but in a few "dark days" the evaporation of water would cease, the atmosphere and the oceans would gradually stop circulating, and the rivers of the earth would empty the water of the continents into the ocean. Photosynthesis could no longer continue and plants would begin to wither. The waters of the oceans would give up their heat of fusion and freeze solid. As the eternal night became colder, all plant and animal life would die. The world would be a changeless, lifeless planet.

23-2 Temperature and Life in the Solar System

Life of the kind we know, and perhaps the only kind possible, is critically dependent upon the presence of water. Not only must water be present, but it must not be permanently below the freezing point (0°C) or above the boiling point (100°C). The temperatures of the surfaces of some of the planets are given in Table 23-1. Only Mars and the earth

TABLE 23-1

Planet	*Temperature*		*Velocity of escape*
Mercury (sunny side)	417°C	783°F	2.2 mi/sec
Venus (sunny side)	430	800	6.3
Earth (average)	14	57	7.0
Mars (warmest part)	12	54	3.1
Jupiter	−138	−216	37.3
Saturn	−153	−243	22.4

have surface temperatures that enable water to maintain a liquid form, and this fact alone seems to limit life as we know it to a small portion of the solar system.

The temperatures at the surface of a planet play an im-

portant part in determining whether or not a planet can retain an atmosphere. The force of gravity at the surface of a planet determines the velocity that an object must achieve before it can escape from the gravitational field of the planet. On the earth, the velocity of escape is approximately 7 mi/sec. A planet's surface temperature is a measure of the average kinetic energy of the molecules of gas at the planet's surface and, therefore, a direct measure of their velocities. If these velocities exceed the velocity of escape, the planet will be unable to maintain an atmosphere. Our moon and the planet Mercury are unable to hold any appreciable atmosphere because of these circumstances. On the larger planets, the force of gravity is strong enough to make the velocity of escape much greater than it is on the earth. With their lower surface temperatures and their greater escape velocities, these planets are able to hold atmospheres composed of gases that are considerably different from those in our own atmosphere. Astronomers have studied these planets by spectroscopic examination and have found this to be true. The earth's atmosphere, then, is in many ways unique in the solar system since its characteristics depend largely upon the peculiar combination of the earth's mass and its surface temperature. In the next few pages we shall examine some of the ways in which this atmosphere is vitally important to our lives.

23-3 The Greenhouse Effect

Suppose we put a thermometer in a sealed, transparent jar and hang a second thermometer outside the jar but near it, as in Fig. 23-2. When the jar and the two thermometers are placed in a darkened room, the temperature inside and outside the jar will be the same. If we place the jar and the thermometer in sunlight, both temperatures will rise. We will find, however, that the temperature inside the jar will rise faster and remain higher than that outside. This phenomenon is known as the *greenhouse effect*, and to explain it we shall refer to Chapter 20 and the discussion of radiation. The light coming from the sun is very complex. It contains electromagnetic energy in a variety of forms, and its spectrum ranges all the way from radio waves to very short ultraviolet waves (Fig. 10-27). In examining the sun's spectrum, however, we must remember that we are at the bottom of an ocean of air that is transparent to many wave lengths but does not allow other wave lengths to penetrate. In fact, the atmosphere is known to be almost opaque to both long infrared waves and short ultraviolet waves.

In Chapter 20 we learned that even when a stove is cold, it will radiate energy in the form of infrared light. The colder the object the longer will be the wave lengths of the energy radiated, and as an object becomes warmer the wave lengths of the radiated energy will become shorter (see Fig. 12-6).

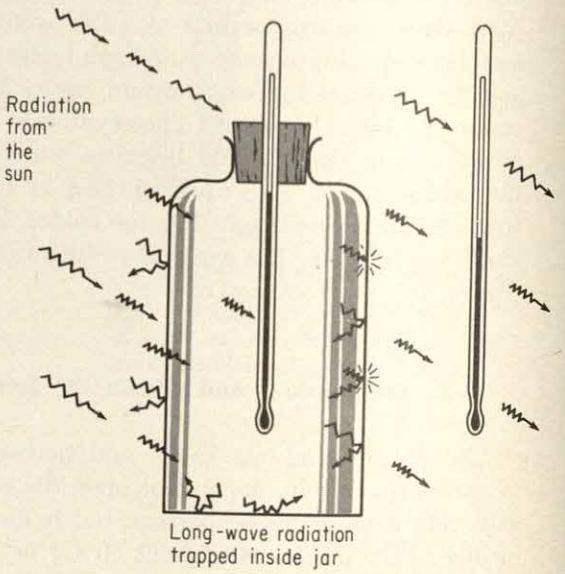


Fig. 23-2. The greenhouse effect.

The earth itself is radiating energy, even though to us it does not seem particularly warm. The lengths of the electromagnetic waves coming from the sun are in general somewhat shorter than those being radiated from the earth. Our main interest, however, is the effect that the atmosphere has on the different forms of radiant energy. The short-wave ultraviolet light finds the air to be opaque, which is fortunate for us, for the light of these wave lengths carries enough energy to be destructive to life on earth. Upon striking the atmosphere, ultraviolet light changes some oxygen molecules, O_2, to oxygen atoms ($O + O$) that can then combine with other oxygen molecules to form ozone molecules, O_3 (see p. 362). The ozone, in turn, will absorb the somewhat longer wave ultraviolet light and change it back to oxygen O_2 molecules. This process ensures a more or less constant balance between the ozone and the oxygen in the atmosphere, a balance that establishes the amount of ultraviolet light that can penetrate the atmosphere.

Most of this combination and recombination of oxygen in the atmosphere takes place in the region sometimes called the *ozonosphere*, located between 15 and 20 miles above the earth's surface (Fig. 23-3). Even there the amount of ozone

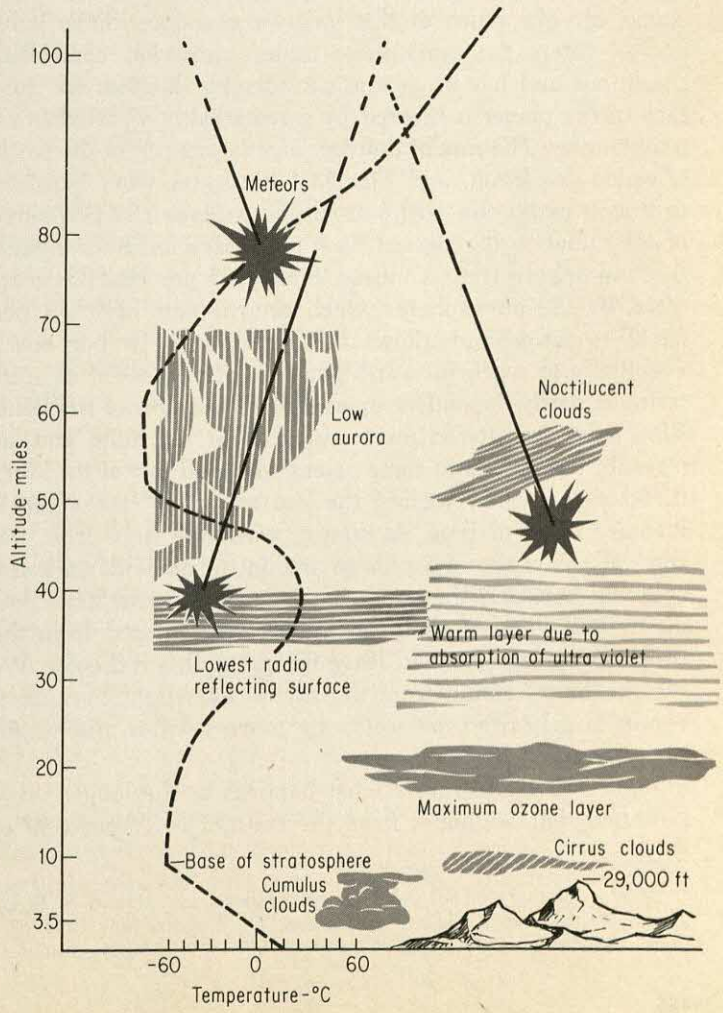


Fig. 23-3. The earth's atmosphere.

is only one part in five million. It is such a small part of the earth's atmosphere, in fact, that if we were able to collect all of it in one layer at the temperature and pressure existing at sea level, it would form a layer only two to three millimeters in thickness, roughly as thick as sixty pages of this book. Nevertheless, this small amount of ozone located high in the atmosphere so effectively absorbs ultraviolet light, which might otherwise be very harmful to life, that practically no ozone is produced in the lower 6 miles of air. An interesting conjecture has been made by scientists as a result of the discovery of this phenomenon. If during the early stages of the formation of the earth there was little oxygen in the atmosphere, a large amount of high-energy ultraviolet light was probably striking the earth's surface, enough possibly to bring about photochemical reactions that do not occur here today. The possibility that some of these reactions took place has been the basis of certain speculations about the origin of life.*

But to get back to the greenhouse effect, both water vapor and ozone serve to absorb energy from the ultraviolet end of the spectrum. The earth's surface absorbs the incoming radiation and then reradiates it at a longer wave length. Again the water vapor, with help from carbon dioxide and ozone, absorbs much of this long-wave energy. Since light energy enters the atmosphere under somewhat controlled conditions and has its escape considerably limited, the surface of the planet is favored by a remarkably well-balanced temperature. The rate of delivery of solar energy to the earth is called *insolation*, and Fig. 23-4 illustrates what happens to it as it enters the earth's atmosphere. Some (33 per cent) of the radiation is reflected back into space much as though the atmosphere were a mirror. Some (25 per cent) is scattered by the atmosphere, which returns part of it (9 per cent) to space and allows the remainder (16 per cent) eventually to reach the earth's surface. The amount of scattering is partly dependent upon the wave length of the light. Blue light is scattered more readily than red light, and, as a result, the blue is to some extent scattered out of the sun's direct rays so that we see the sun as yellow, the complementary color of blue. At sunset, when the light from the sun must pass through a large amount of air with its many colloidal particles, the scattering may be so great that blue, green, and some of the yellows may be scattered from the direct rays of the sun and leave the sun with a red color. We see the light coming from the sky as blue light because it results largely from the scattering process within the atmosphere.

Figure 23-4 also shows what happens to the longer wave radiation that originates from the earth. The 27 per cent of

* For a summary of several such speculations, see Harold S. Blum, *Time's Arrow and Evolution* (Princeton, N. J.: Princeton University Press, 1951).

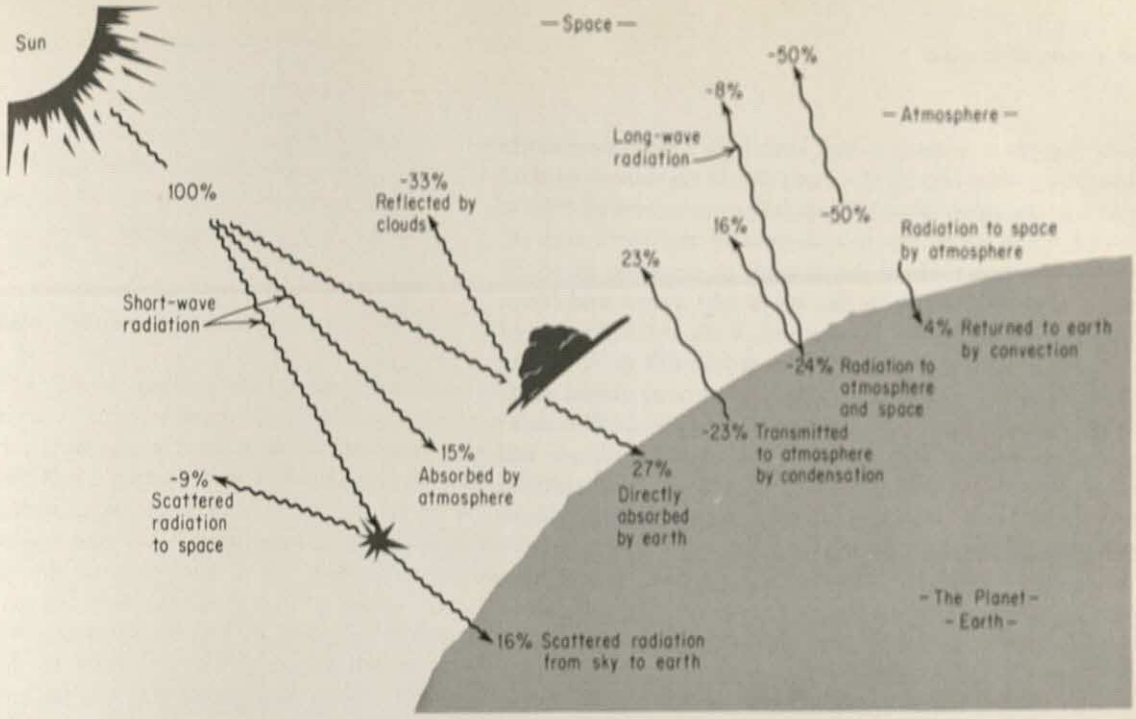


Fig. 23-4. The radiation balance of the planet earth.

the insolation that is directly absorbed by the earth, and the 16 per cent that comes to the ground from scattering tend to raise the temperature of the earth's surface. Of course, the percentages given here are approximations and are being refined as we learn more about the atmosphere. Figure 23-4 does not take into account the radiant energy from the sun that is stored by the plants through the photosynthetic process. Although to us this may seem to be a large oversight (and does indeed represent some 1.8×10^{21} calories per year), the sunlight that falls upon the earth during this same time is 1.27×10^{24} calories, so that the amount used in photosynthesis represents only 0.14 per cent of the sunlight available. In order for the temperature of the earth to remain constant, radiation must remove from the earth an amount of energy equivalent to that received. As we can see, some energy is radiated directly to space while some is absorbed by the atmosphere. When the energies in Fig. 23-4 are all totaled, however, we see that the amount of energy that leaves the planet is equivalent to the energy that arrives from space.

23-4 The Greenhouse Effect and the Glacial Ages

The relationship between heat gain and heat loss by the earth may not always have been so closely balanced, and the amount of heat coming from the sun may actually have varied through the geological ages. Several hypotheses may illustrate this point. Suppose that the solar system, in its travels through the universe, passed through some dust

clouds capable of scattering and absorbing light. The amount of insolation absorbed by the earth would be decreased, and for a time the earth would be radiating more energy than it received. Only when the temperature of the earth was reduced to the point where the amount of radiation given off equaled that received from the sun would a new and lower temperature equilibrium be reached. With the reduction of the earth's temperature, more snow would fall each winter, less would melt in the summer, and snow and glacial fields would gradually build up. The snow would make the earth a better reflector of the sun's rays and thereby reduce still further the amount of light energy that it would absorb. In this way, a small reduction of solar radiation might have caused the glacial ages on the earth.

A second way the ice ages might have been produced is through volcanic activity. In 1883, the volcano Krakatoa, located near the island of Java in Indonesia, erupted with such violence that it blew 4 cubic miles of rocks, dust, and soil high into the air. Enough matter was blown into the stratosphere to color the sunsets of the earth for the next three years. At that time there were not enough laboratories equipped to measure the amount of solar radiation being received from the sun to give us a conclusive measure of the reduction in insolation. Nevertheless, it is agreed that there was a reduction in energy being received from space, and it is thought by some scientists that an increase in the volcanic activity of the earth would put enough dust in the earth's atmosphere to produce another glacial age.

Another way that a glacial age may have come about is through an increase in the radiation retained by the earth. Volcanoes, in addition to throwing up dust and ash, could increase the amount of carbon dioxide in the atmosphere to the point where it would block much of the present long-wave radiation from leaving the earth. As a result of the increased energy retained by the earth (incoming short-wave rays plus additional long-wave radiation from the earth returning to it), more water would be evaporated and more clouds would form. Over the colder regions of the earth, the water vapor would condense into rain and snow, and with more of the earth covered by ice and snow (which have a greater ability to reflect light than the darker colored earth), the sun's energy could not be utilized as well and the snow caps of the earth would grow. Only when the temperature, especially in the tropics, was reduced to a point where the amount of evaporation was substantially decreased would the snow packs begin to decrease and the glacial age come to an end.

It may seem surprising that a glacial age could be brought about by either an increase or a decrease in insolation or by an increase in the internal heat content of the earth with a resulting step-up of volcanic activity. Of course, if the amount of energy received from the sun were greatly increased, there would be no glacier formation because it

would get too hot on the earth's surface. Fortunately, the heat being radiated from the sun is remarkably constant; it seems to have varied by less than 2 per cent during all the years it has been studied.

23-5 Water in the Air

The lowest part of the atmosphere is called the *troposphere*. It extends from the earth's surface to a height that varies from about 5 miles above the poles to about 11 miles above the equator. Over the continental United States it is about 7 miles high. This is the region of clouds, changing weather, and a temperature that drops with increasing altitude. Marking the top of the troposphere is the *tropopause* where the temperature ceases to drop.

The *stratosphere* extends from the tropopause to a height of about 50 miles. In this region the air movement is horizontal with winds strong and steady. The air is clear, practically without dust or water vapor. At the base of the stratosphere the temperature is likely to be about −67°F in the middle latitudes. At higher elevations the temperature rises to the neighborhood of 90°F and then drops again to well below freezing at the top of the stratosphere near 50 miles. Above this is a region 500 miles thick of very thin air where large numbers of electrically charged particles (ions) give it the name *ionosphere*. This zone contains much of the aurora activity and the layers that reflect the ordinary radio waves back to earth. The outermost layer of the atmosphere, the *exosphere*, extends upward with decreasing density to the near reaches of outer space. It is only in the troposphere that water vapor seems to play an important role.

The air around us is practically never, if ever, perfectly dry. On the other hand, it is not as wet as some of us might think, for the water vapor that is one of the most important constituents of the atmosphere seldom makes up as much as 4 per cent by volume of the atmosphere. Yet from this small fraction of the atmospheric components comes all the water that powers our hydroelectric projects, waters our fields, and brings our refreshing rains and winter's snows. Our very lives depend upon this 4 per cent of the atmosphere, and yet 90 per cent of the water in the air is present in the lower 25,000 feet. Figure 23-5 summarizes this water movement into and out of the atmosphere.

The mass of water per unit volume of moist air is called the *absolute humidity* of the air, or, in other words, the amount of moisture contained in the air. It can be measured in grams of water per cubic meter or in grains of water per cubic foot. The amount of moisture that the air can hold, largely determined by the temperature, is called the *saturation humidity*. If we divide the absolute humidity by the saturation humidity and change the resulting decimal to per cent, we have a value that is called the *relative humidity*.

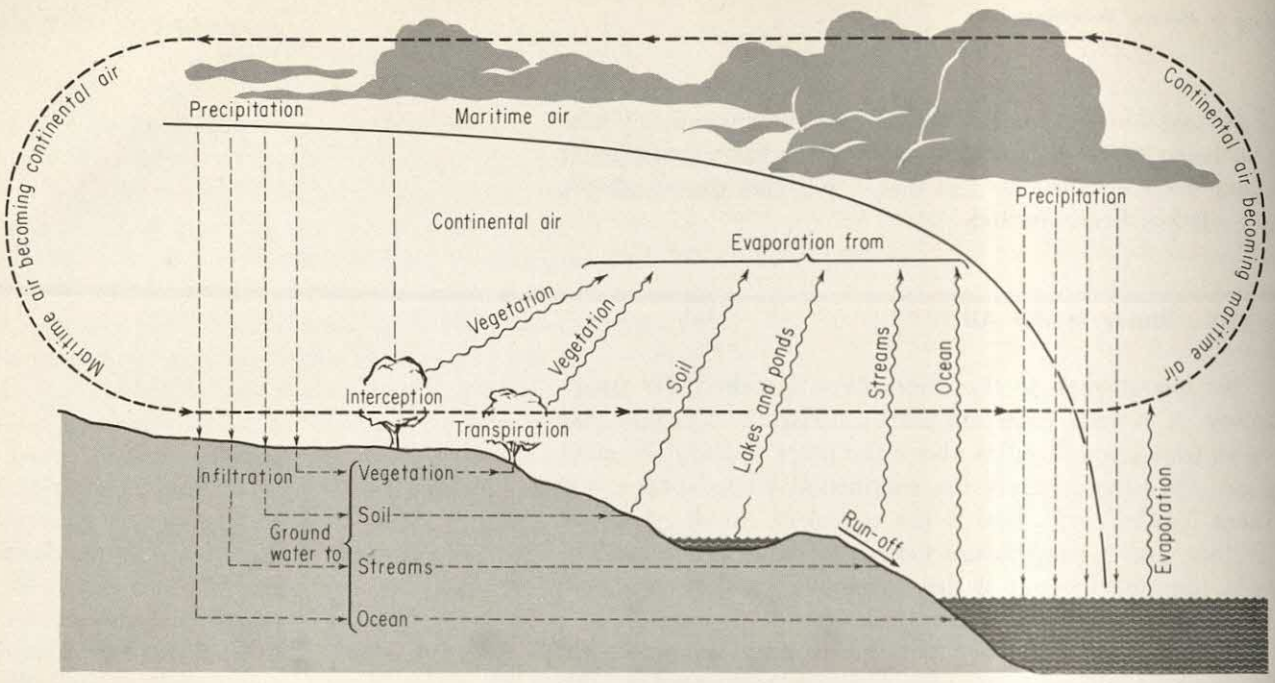


Fig. 23-5. The water cycle: a summary of the processes by which water vapor enters the atmosphere.

Or, more simply, relative humidity is the ratio of the quantity of water vapor actually present in any volume of air to the quantity required to saturate the same volume of air at the same temperature. In formula form:

$$\text{relative humidity in per cent} = \frac{\text{wt of water in air}}{\text{wt of water air could hold}} \times 100 = \frac{\text{absolute humidity}}{\text{saturation humidity}} \times 100 \qquad (23.1)$$

We can see that the relative humidity for saturated air must be 100 per cent while for air that is perfectly dry it is 0 per cent.

23-6 The Measurement of Relative Humidity

There are a number of ways in which we can determine relative humidity, but we shall limit ourselves to two of these methods. If we put ice in a glass of water and let the glass cool down, the glass may reach a temperature at which the air surrounding it will be saturated with moisture. With a lowering of temperature, this air is unable to hold any more water vapor, and some moisture will begin to form on the glass which will "cloud over" or even "sweat." This is the way in which dew forms, and the temperature at which moisture just condenses out of the air is called the *dew point.* In the case of our glass with ice water in it, the temperature of the water in the glass is for all practical purposes equal to the temperature of the glass. The air right at the surface

of the glass is also practically at the temperature of the glass and the water, and, as a result, we have an easy way of determining with reasonable accuracy the dew point temperature. If the air is saturated, any decrease in air temperature will cause moisture to condense, and the dew point and the air temperature will be the same. As the amount of moisture in the air decreases, the temperature will have to be lowered more and more in order for dew to form. Table 23-2 gives us the information we need to determine the relative humidity.

TABLE 23-2 WEIGHT OF WATER VAPOR IN 1 CU M OF SATURATED AIR

Temperature	*Water vapor*
0°C or 32°F	4.8 g
5°C or 41°F	6.8 g
10°C or 50°F	9.3 g
15°C or 59°F	12.7 g
20°C or 68°F	17.1 g
25°C or 77°F	22.8 g
30°C or 86°F	30.0 g
35°C or 95°F	39.7 g

An example will clarify the use of this table. If air in the atmosphere has a dew point of 15°C and a temperature of 20°C, what is the relative humidity? The dew point tells us the temperature at which the air is saturated (15°C), the table tells how much moisture the air has (12.7 g/m³), and the air temperature determines how much water the air could hold (17.1 g/m³ at 20°C). The relative humidity, therefore, is:

$$\frac{\text{wt of } H_2O \text{ in air}}{\text{wt of } H_2O \text{ air could hold}} \times 100$$

$$= \frac{12.7 \text{ g/m}^3}{17.1 \text{ g/m}^3} \times 100$$

$$= 74.2 \text{ per cent}$$

The air has a relative humidity of 74.2 per cent, which means that the air contains 74.2 per cent of the water vapor it could hold if it were saturated.

Suppose we ask a further question: How much moisture would condense per cubic meter if the air in the previous example were cooled to a temperature of 10°C? The dew point tells how much moisture the air has (12.7 g/m³ or 74.2 per cent of 17.1 g/m³ = 12.7 g/m³). But at 10°C, it can hold only 9.3 g/m³, and the difference must condense: 12.7 − 9.3 = 3.4 g/m³. For each cubic meter of the air that cools from 20°C to 10°C, 3.4 g of water will condense. If we knew how many cubic meters were in a cloud formed under these conditions, we would have a very good estimate of how much rain might fall.

A second method by which we can find the relative humidity is to use wet- and dry-bulb thermometers. If we cover a thermometer bulb with a thin wet cloth and circulate air around it, the drier the air that is circulated, the more rapid will be the evaporation of the water from the wet cloth. We learned in Chapter 22 that whenever water is vaporized it absorbs an amount of heat called the heat of vaporization. The more rapid the evaporation caused by the dry air, the more heat is abstracted by the evaporating water from the mercury and the glass of the thermometer, and the lower will be the temperature reading. By knowing the temperature of the air, the dry-bulb thermometer reading, and the wet-bulb depression (the number of degrees difference between the dry bulb and the wet bulb), we can determine the relative humidity by referring to tables published by the U.S. Weather Bureau. An abbreviated version of such data is given in Table 23-3.

TABLE 23-3 **RELATIVE HUMIDITY**

Dry-bulb thermometer (°F)	*Difference between wet- and dry-bulb thermometers in °F*																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
60	94	89	83	78	73	68	63	58	53	48	43	39	34	30	26	21	17	13	9					
62	94	89	84	79	74	69	64	59	54	50	45	41	36	32	28	24	20	16	12	8	4			
64	95	90	84	79	74	70	65	60	56	51	47	43	38	34	30	26	22	18	15	11	7	4		
66	95	90	85	80	75	71	66	61	57	53	48	44	40	36	32	29	25	21	17	14	10	7	3	
68	95	90	85	80	76	71	67	62	58	54	50	46	42	38	34	31	25	23	20	16	13	10	6	3
70	95	90	86	81	77	72	68	64	59	55	51	48	44	40	36	33	29	25	22	19	15	12	9	6
72	95	91	86	82	77	73	69	65	61	57	53	49	45	42	38	34	31	28	24	21	18	15	12	9
74	95	91	86	82	78	74	69	65	61	58	54	50	47	43	39	36	33	29	26	23	20	17	14	11
76	96	91	87	82	78	74	70	66	62	59	55	51	48	44	41	38	34	31	28	25	22	19	16	13
78	96	91	87	83	79	75	71	67	63	60	56	53	49	46	43	39	36	33	30	27	24	21	18	16
80	96	91	87	83	79	75	72	68	64	61	57	54	50	47	44	41	38	35	32	29	26	23	20	18
82	97	92	88	84	80	76	72	69	64	61	58	55	51	48	45	42	39	36	33	30	27	25	22	20
84	97	92	88	84	80	76	73	69	65	62	59	56	52	49	45	43	39	37	34	32	28	26	24	21

To understand the importance of humidity and the moisture that is condensed from an air mass, we must realize that virtually all the water that we use in our lives comes to us by way of rain. And the ultimate source of all rain is, of course, the ocean. Some water is recycled by evaporation from lakes, rivers, plants, and moist soil, but all moisture first comes from the ocean and eventually ends there. It is estimated that 16 million tons of water are evaporated from the earth's surface each second, and, thus, an average of 16 million tons must rain each second. Not only enormous quantities of water but also enormous quantities of heat are involved in the process of rainfall. For instance, if one inch of rain falls on a square mile of our earth, the heat given off by the condensation of this amount of water is equal to 1.5×10^{11} Btu, which is equivalent to the heat produced in the burning of 6,400 tons of coal. Worldwide, the condensation of rain forms a gigantic heat exchange process. The heat given off each second over the entire earth by condensing rain is roughly 32×10^{12} Btu or the equivalent of burning

each second 1,600,000 tons of coal. This rate could be matched only by burning up the whole annual production of coal in the United States in four minutes.

23-7 Condensation in Rising Air Masses

Most of us erroneously think that rain brings cooler weather, when it is usually the other way around; cooler air brings with it rain because it condenses moisture out of the air. This we know although we are not often in a position to see what happens in the cloud where the condensation of the rain takes place.

Let us test our thinking about heat being given off when moisture is condensed by taking a look at the temperature of air as it rises up a mountainside and forms a cloud, a phenomenon that occurs regularly in the mountainous regions of the world (Fig. 23-6). In the Cascade Mountains, as an example, the wind usually blows from the west and brings considerable moisture from the Pacific Ocean. Suppose the sea level temperature is 70°F and the dew point is 52°F. As the altitude increases, the temperature becomes cooler, and at 1000 feet it is 64.5°F and the dew point is 51°F. At 2000 feet, they are 59°F and 50°F respectively. For every rise of 1000 feet in the clear air of our example, the temperature drops 5.5°F. This drop in temperature with an increase in elevation is called the *lapse rate.* In this case, where the drop in temperature takes place in clear dry air it is called the *dry adiabatic lapse rate.* The term *adiabatic* simply means that no heat is added from sources outside the system being considered—in our example, this means that all the heat involved comes from the air mass itself. Figure 23-6 shows the temperatures and dew points at succeeding elevations. Note that the dew point drops 1°F for every 1000 feet of rise in elevation.

At 4000 feet, the air temperature and the dew point are both 48°F. The air under these conditions is saturated, and any further drop in temperature will cause condensation and

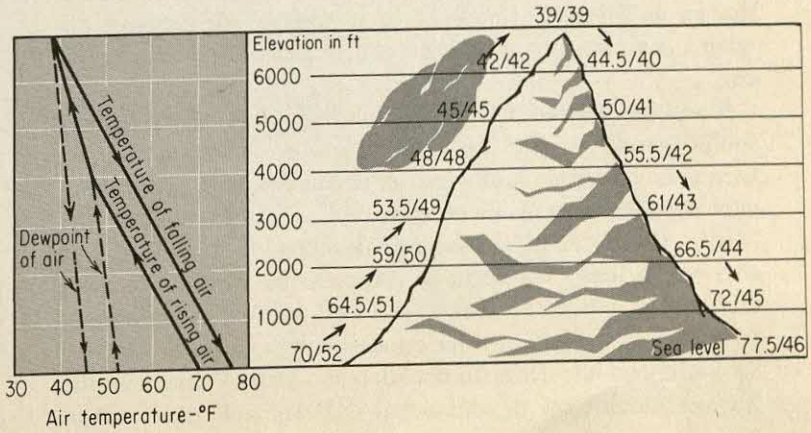


Fig. 23-6. Air rising over a mountain. The two figures separated by the slant line (/) are the air temperature (°F) and the dew point (°F), respectively.

clouds to form. In the next 1000 feet, the temperature of the foggy air or cloud drops only 3°F, not as much as previously. Before you read further see if you can tell why the air temperature decreases only 3°F/1000 ft in the cloud whereas it decreases 5.5°F/1000 ft in dry air. We already have the clue to the answer, for we know that whenever moisture condenses it gives off heat. As the temperature drops below the dew point, more and more moisture condenses and more and more heat is added to the air mass that is rising. The temperature drops as the elevation increases but not as rapidly as it would if moisture did not condense from the air mass. We call the change in temperature per 1000 feet in air in which moisture is condensing the *moist adiabatic lapse rate.* In air that is saturated, the dew point and the air temperature are the same, and, as a result, the dew point in our example drops at the same rate as the moist adiabatic lapse rate above the elevation where clouds are forming, so the air is full of moisture at the mountain top.

Figure 23-6 shows that the air crosses the top of our mountain at an elevation of 7000 feet. As the air starts to drop in elevation on the back side of the mountain, it is warmed because its density increases (as a gas compresses, its temperature increases). The moment the air temperature increases, it can hold more moisture, and so it stops condensing moisture and becomes clear again. The "dry" air warms up at the same rate that it cooled off. The dew point, too, will resume its change at the same rate it maintained in the dry adiabatic lapse range. When the air returns to sea level again it is warmer, having a temperature of 77.5°F as compared to 70°F; the dew point also is lower, 46°F compared to 52°F. The air is warmer because it is drier. The hot, dry wind that warms the lee side of a mountain and melts its snow is known as a chinook in the western United States. The warmer temperatures and reduced rainfall on the lee side of the mountain are responsible for the differences in vegetation found there.

With this background, we can now generalize and say that the rainy regions of the earth will be the regions where air often rises in large quantities. We can also see that where the air is being warmed, as in regions of descending air where compression occurs, no precipitation is likely to occur.

Since the United States lies in the region of prevailing southwesterly winds, air usually moves from the Pacific Ocean inland from southwest to northeast. This process occurs day after day as air crosses the Cascade Mountains, the Sierras, the Rockies, and the Appalachians. In each case, but to a much lesser extent in the Appalachians, where coastal winds may modify the effect, the west side of the mountain range is moist and rainy and the east side is dry and warmer. The arid regions of California such as Death Valley and the Mojave Desert; the dry area that is being put under irriga-

tion in the Columbia Basin; the desert climate of Nevada, Utah, and Arizona; the drought areas of our great plains—all these stem from the effects of mountains on the temperature and moisture of air.

23-8 Air Masses: Their Source and Movement

An air mass is a large section of the troposphere in which the temperature and humidity are fairly constant at a given level. These air masses owe their characteristics to the surfaces over which they originate. If they are *tropical,* they are abbreviated with a T; if *polar,* with a P. When the source region is the extremely cold, ice-covered Arctic, the air mass may be *Arctic,* A, whereas *Equatorial,* E, refers to the moist, hot air along the equator that does not directly affect the United States. Occasionally a hot, dry air descends from

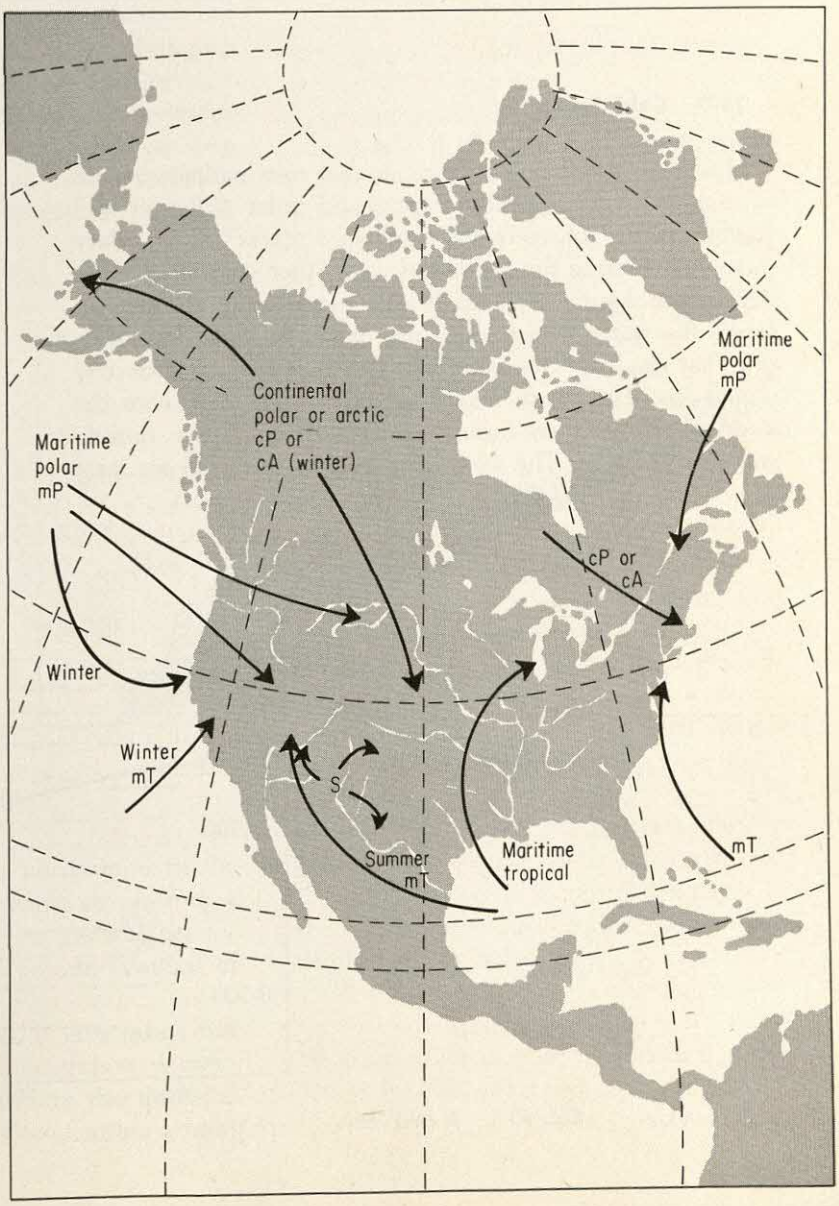


Fig. 23-7. Major air masses which affect North America.

the middle and upper troposphere to parch and bake our Southwest; this air mass is known as *Superior*, S.

In addition to the notation given, the source region is either an ocean or a continent. Air masses that originate over an ocean are called *maritime*, m, and are more humid than those originating over the land, called *continental*, c. Table 23-4 summarizes the characteristics of the four major air masses that generally affect North America; Fig. 23-7 shows the general areas of their influence.

TABLE 23-4 AIR MASSES AND THEIR MAJOR CHARACTERISTICS

Name	*Abbreviation*	*Characteristics*
Continental polar	cP	dry, cold
Maritime polar	mP	moist, cold
Maritime tropical	mT	moist, warm
Continental tropical	cT	dry, warm

23-9 Cold Fronts

Most of our rain comes from air that rises without benefit of mountains. Periodically, continental polar (cP) air will flow out of the source region southward across the boundary states and then to the south and east. Since cold air is more dense than warm, moist air, it stays close to the ground and forces the warm, moist air to rise out of its way. The situation that results is comparable to that in which air rises over mountains. The warm, moist air cools until it reaches the dew point and then forms clouds and eventually rain or snow (Fig. 23-8). The advancing edge of the cold air mass

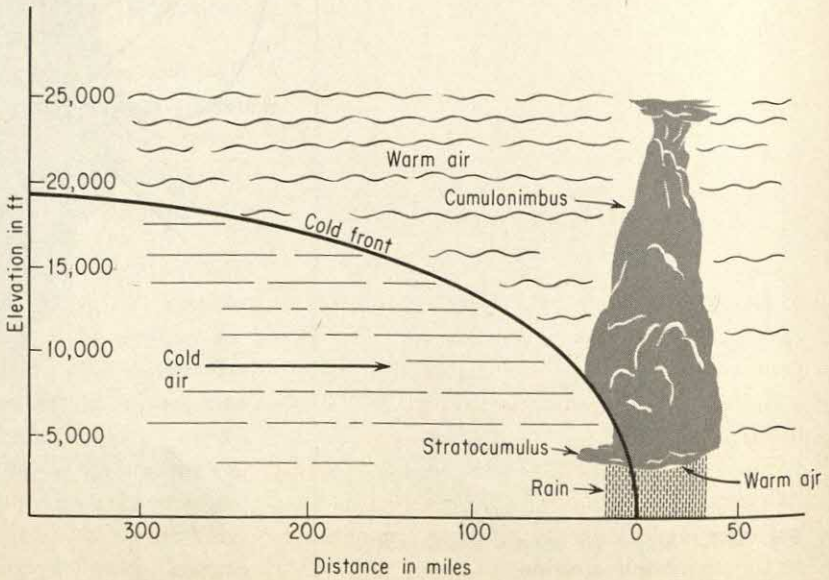


Fig. 23-8. A cold front.

is called a *cold front*, forming one of our most recognizable weather phenomena.

The passage of a cold front is often marked by squalls, thundershowers, quick storms, and a drop in temperature. On the ground, the barometer shows a rise in pressure, testifying to the greater density of the cool air above. Since the cold air is more dense than the warm air and is forcing its way along the ground, the rise of the warm air is quite abrupt. As a result, the cold front and its associated storms, clouds, and precipitation usually covers a relatively short distance, and the weather is apt to clear up quickly after it passes.

23-10 Warm Fronts

As we have seen, the tropical Atlantic Ocean around Bermuda produces a warm, moist air mass known as maritime tropical (mT). This air periodically moves across the Gulf of Mexico and up the Mississippi Valley, bringing warm, humid weather to the central states. The forward edge of such an air mass is called a *warm front*. As the warm air moves northward and eastward, it must force its way through, or push aside, the cold air that blocks its path (Fig. 23-9). Since it is less dense than the cold air, it tends

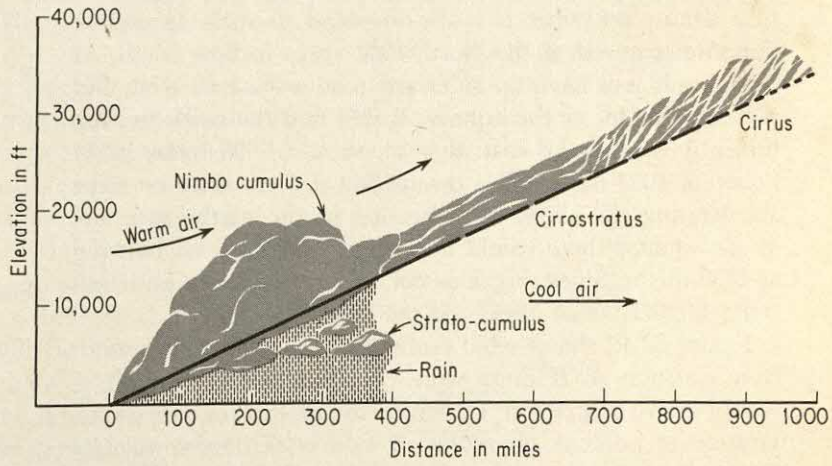


Fig. 23-9. A warm front.

to rise above the cold air, though not as abruptly as in the case of an advancing cold front. The warm air rises in a gentle slope so that the condensation affects the weather for a thousand miles or more ahead of the ground position of the front.

The approach of a warm front is usually very rainy, but its passage is accompanied by a rise in temperature, a gradual clearing of the weather, and a change in the direction of the wind. Even when the rain stops, the air often retains

a fairly high humidity. There are, of course, all types of variations of cold fronts and warm fronts, and when the two fronts collide a cyclonic storm may develop that will cover hundreds of square miles of the earth's surface.

23-11 Circulation of the Atmosphere and the Coriolis Effect

The movements of the warm fronts and cold fronts result from the large circulatory movements of the earth's atmosphere. This movement of air (or wind) depends in large measure on temperature differences as its source of energy. These temperature differences in turn are often caused by heat gained or released in the condensation processes we have just been considering.

Figure 23-10 indicates that there is an excess of radiant heat at the equator and a deficit above latitude 37°. At the equator the warm air rises and at the poles the cooled air flows groundward. From this we might assume that there are large-scale convection currents set up over the earth such as those shown in Fig. 23-11, but, although this pattern is basically true, the situation is not this simple. We shall have to understand something about the Coriolis effect before we can see how the United States happens to be in a region of prevailing southwest winds.

We seldom think there is anything strange about water or air flowing north or south, but, taking an exaggerated case, let us see what is really involved in such an action. Suppose some air at the North Pole starts to flow south. At the start it will have no sideward motion east or west, but if it flows as far as the equator it will find the earth turning under it toward the east at a speed of 24,800 miles in 24 hours, or 1033 mi/hr. This means that if there were no force accelerating the air in the direction of the earth's rotation, at the equator there would be a wind from the east blowing at 1033 mi/hr. Since this does not occur, there must be modifying factors which affect the motion of the air.

Figure 23-12 shows what really happens as air flows south from position A. If there were no rotation of the earth, it would arrive at point B. However, the earth's rotation at the latitude of point A gives the air two velocities—a velocity V_s southward and a velocity V_e eastward. The resultant of these two velocities is the true velocity of the flowing air and is indicated by V_t. We might think that the wind at B would be from the northwest because the final velocity of the southward flowing air has an easterly component, but it does not, for the earth too has been turning. While point A was moving to point A' on Fig. 23-13, point B was moving to point B'. The southward flowing air maintains its eastward component and flows south and east, but it does not move east as fast as the earth at point B, so, lagging behind, it arrives at point A''. Notice that point B is directly south

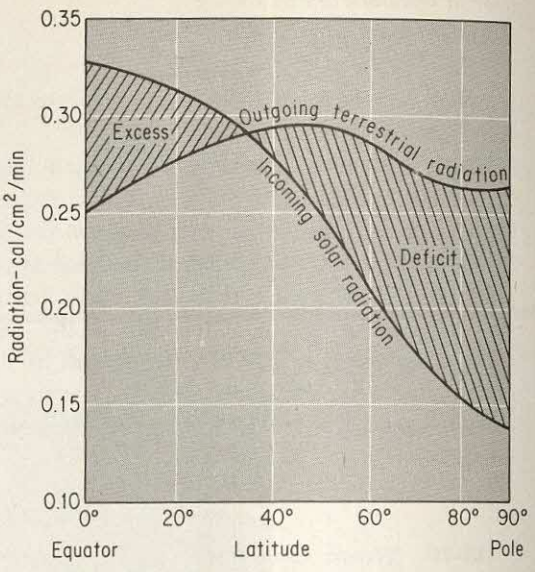


Fig. 23-10. The incoming and outgoing radiation at each latitude.

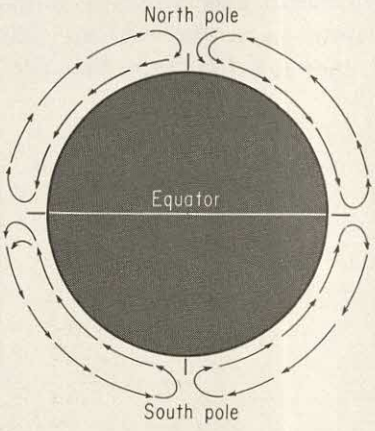


Fig. 23-11. An idealized convection current caused by the unequal heating of the earth.

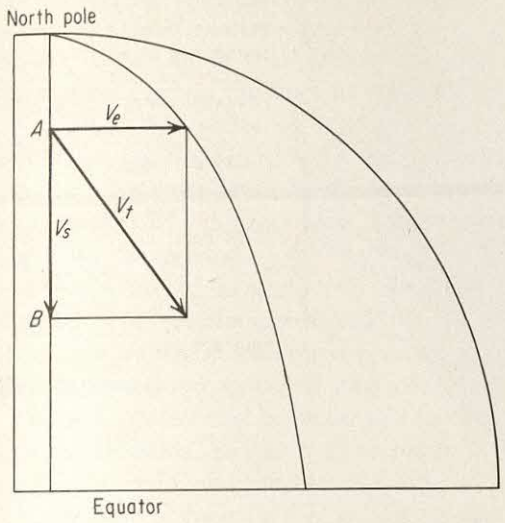


Fig. 23-12. Air moving from north to south in the Northern Hemisphere. The earth's motion causes this air to have two velocities—one southward, V_s, and one eastward, V_e. The resultant velocity is the vector addition of these two velocities, V_t.

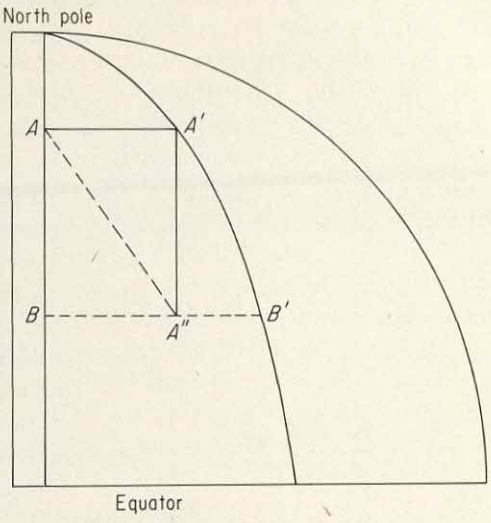


Fig. 23-13. While the air was moving south from A to A'', point A had moved to A', and point B had turned from B to B'.

of point A, and B' is south of A'. So to a person at B or B' the wind is blowing from the northeast. Notice that in the Northern Hemisphere the deflection of the wind has been to the right.

Suppose now that some air is flowing from the equator north to latitude 30°. If at the equator the air is still, it must, of course, be moving just as fast as the earth at that point—1033 mi/hr. At latitude 30°, the earth's surface has an eastward movement of approximately 894 mi/hr. If we neglect the friction of the earth and other modifying factors, this means that the speed of the air at the equator would be 1033 — 894 or 139 mi/hr greater than at latitude 30°. As the equatorial air flows northward, it will be going faster than the earth at latitude 30° and will blow not from the south but from the southwest (Fig. 23-14). Again we notice that in the Northern Hemisphere the deflection of the wind due to the earth's rotation is to the right, which is always the case in the northern half of the earth. In the Southern Hemisphere, the air is deflected to the left. This deflection is known as the *Coriolis effect* and is, of course, dependent upon the latitude and the velocity of the wind.

As air flows south from the north polar region, it deflects to the right and produces the northeast winds found in the Arctic. Approaching the equator, it results in the northeast trade winds, which helped bring Columbus across the Atlantic (Fig. 23-15). With our knowledge of the modifying effect of the earth's rotation on air movements, we can now

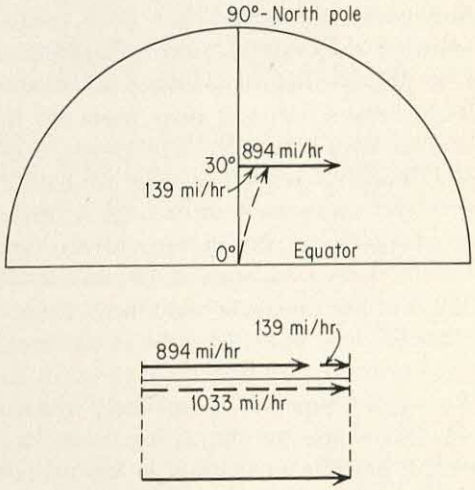


Fig. 23-14. The relative velocity of still air at the equator and at latitude 30°.

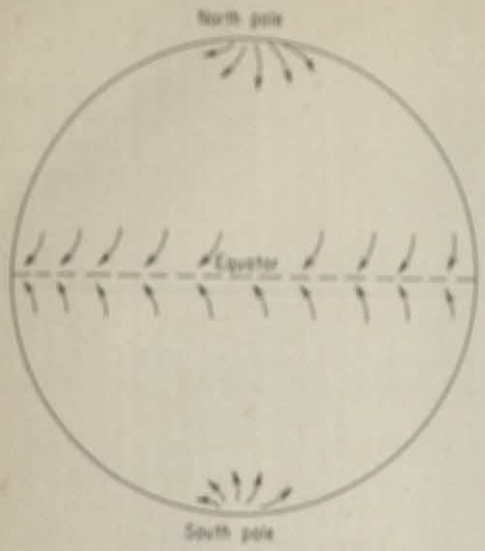


Fig. 23-15. The Coriolis effect in the Northern and Southern Hemispheres.

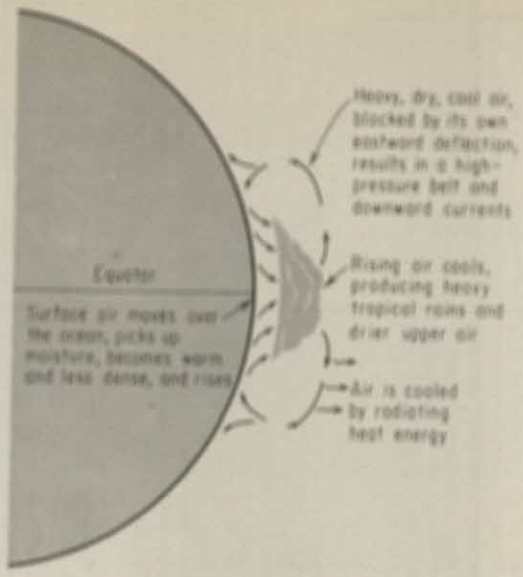


Fig. 23-16. The formation of the subtropical high-pressure belt.

understand why the idealized atmospheric convection currents shown in Fig. 23-11 differ from what really happens.

In the Northern Hemisphere, as the high altitude winds flow north from the equator they are deflected more and more to the right until in the vicinity of 30° to 35° north latitude, they flow so much to the east that in comparison to their northward movement there is a partial damming or blocking of further northward flow. As these winds pile up they create a high-pressure region known as the subtropical high-pressure belt, which is a band containing descending air currents that extend from high elevations to the earth's surface. When these descending air currents reach the surface, they move outward away from this high-pressure belt.

Another reason for this high-pressure belt in the region of latitude 30° is related to what we have learned about air density and water vapor in the air. As the air sweeps southward along the earth's surface to the equator, more than half of its flows over open water, so that in addition to becoming warmer the air is moistened. These two factors work together to cause the air to rise at the equator (Fig. 23-16). But as we have seen, the air cools as it rises and thus can hold a smaller amount of moisture, which causes it to condense and release the drenching rains that characterize the tropical zone. The air is dried in this process, and as it flows northward at high elevations it loses heat through radiation. Since it is removed from any source of water vapor, it remains dry. This cool dry air has a high density and as it approaches the region of 30° latitude, its temperature, density, and eastward deflection join to produce a high-pressure belt. This being the case, we would expect the areas around latitude 30° to be relatively hot and dry, and this prediction of our theoretical analysis is borne out by

the observations we can make on the earth. As the air flows outward from the subtropical high-pressure region, some of it flows north along the earth's surface and is deflected to the right or east by the Coriolis effect. This produces a southwest wind that shifts as it moves northward to a west-southwest wind (Fig. 23-17). In the region of 60° to 65° latitude, these winds are known as the *prevailing westerlies*, and the air in them is warmed and moistened by its association with the ground.

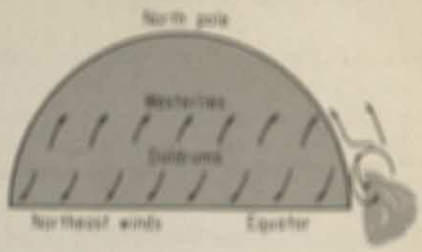


Fig. 23-17. The formation of prevailing westerlies.

The downward movement of air at latitude 30° does not dispose of all of the northward flowing high-level air, and there is a sufficient poleward pressure gradient to continue the northward movement of some of this air, which is further dried and cooled and becomes still more dense. At the pole, this air descends, and as it flows south, it is deflected to the right in a cold, dry northeast wind. In the region of 60° latitude, the surface air that is flowing northward meets this southward flowing polar air. The warmer moist air from the south rises above the colder more dense air from the north, and in rising is cooled by expansion so that it loses moisture by precipitation (Fig. 23-18). These rising surface winds join the northward flowing high-level air and go on to pile up at the poles.

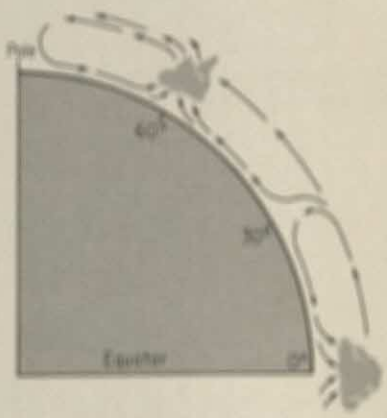


Fig. 23-18. The accumulation of air at high latitudes results in a high-pressure area at the pole.

In the grand cycle of the earth's air currents, air leaves the tropical regions and accumulates at the polar regions. North of 60° to 65° latitude, air flows steadily southward along the surface, producing a front of cold air known as the *polar front*. Periodically this polar air breaks out of its confines and flows southward, bringing cold, dry (continental polar) air into the northern states of this country. Sometimes the flow of cold air is so large that it extends as far south as the Gulf states. These outbreaks of polar air tend to restore equilibrium with the equatorial air by returning some of the excess of winds that have flowed up to the North pole. The circulation of the atmosphere moves air masses, which are warm or cold, moist or dry depending on the location of their source region, across the earth and in large measure determines the climate that exists in any given area.

As we consider this flow of air that moves back and forth across the earth, we can see that fundamentally all air movement results from heat differences. It is in the process of exchanging heat from places of excess to places of deficit that the atmosphere is churned up into great continent-spanning currents of air, which in turn bring us our day-to-day weather—our wind, rain, hail, sleet, and snow. Movements of the atmosphere also dominate the ocean currents that aid the air in transporting large quantities of heat all around our earth. Life, in fact, would be impossible without these movements of the atmosphere that equalize the temperatures of our planet and make large areas of it habitable.

33-11 Circulation of the Oceans

Although water forms only 0.025 per cent of the mass of the earth, it covers approximately 70.8 per cent of the earth's surface. Consequently, the major part of the short-wave energy that finally arrives at the earth's surface is absorbed by water. Since sea water accounts for more than 96 per cent of the mass of water on the earth, it follows that the seas and oceans of our planet absorb the greatest part of the sun's heat that comes to it. Thus, the circulation of the waters of the oceans is the largest mechanism on earth for the exchange of energy.

As we should expect, the motion of the oceans is greatly affected by the circulation of the atmosphere above it. Since the winds at the equator blow from east to west, it is not surprising to find that there are two broad streams of water in equatorial regions that also run from east to west—the north equatorial current above the equator and the south equatorial current below (Fig. 33-10). Between them is a narrow equatorial countercurrent flowing to the east. The two equatorial currents are deflected by the land masses—and partly too by the rotation of the earth just as in the case of air currents. In the Northern Hemisphere, this deflection is to the right or northward, whereas in the Southern Hemisphere it is to the left or southward. Northern Hemisphere currents flow first to the north and turn gradually eastward and then southward where they again join the equatorial current and complete the cycle.

Fig. 33-10. Circulation of the surface waters of the world's oceans.

33-12 Circulation of the Atlantic Ocean

To gain a better understanding of the currents in the oceans, let us consider the flow of water in the Atlantic Ocean. The northward flowing Gulf Stream is the most

prominent feature of the ocean. The water in this stream comes from the north equatorial current and from a branch of the south equatorial current which debouches off the northern coast of South America and carries some 100 million cubic feet of water per second into the Caribbean Sea from the South Atlantic. These waters flow through the Caribbean Sea and the Yucatan Channel and gather speed as they pass through the straits of Florida. In fact, it is estimated that the water moves at a rate of about 4 mi/hr in a current approximately 50 miles wide and 2 miles deep that transports something like 155 million cubic feet of water each second. This "ocean river" contains 1000 times as much water as the Mississippi normally carries. The warm waters of the Gulf Stream transport tremendous amounts of heat. The surface temperature of the Gulf Stream is approximately 80°F during the first 400 miles of its journey. A cubic mile of water in cooling 20°F will deliver 334,000,000,000,000 Btu or, in more meaningful figures, an amount equivalent to the heat given off in the burning of 7 million tons of coal.

We now have some basis for estimating the tremendous amount of heat that are continuously being transported by our ocean currents. We have seen that water is a particularly good transportation agent for heat because of its high specific heat. You will recall that 1 cu ft of water in cooling 1°F will give off enough heat to raise the temperature of 3000 cu ft of air 1°F. No wonder then that as this warm water flows eastward it warms immense volumes of air. Since the United States is in a region where the atmosphere generally moves from west to east, very little of the air warmed by the Gulf Stream reaches our shores. Our weather is affected, though not as much as we might expect, but Gulf Stream air plays a very important part in modifying the weather of Europe. For demonstrable proof of this, just contrast the habitability of Greenland or Baffin Land with that of Norway, which is at the same latitude.

No story of the circulation of the waters of the Atlantic would be complete without considering how water is returned to the South Atlantic to replace the 100 million cubic feet per second that enter the Gulf Stream. How is this done? Some water flows down to the equator along the shores of North Africa. The rest sinks in the Arctic regions and sinks below the surface, where, much less affected by the surface winds, it flows southward to the South Atlantic.

Besides a lowering of temperature, there is a second way that water descends to an ocean. This second process is best illustrated by what happens in the Mediterranean Sea. The Mediterranean is relatively shallow and evaporation from it is rapid. As the surface waters are evaporated, salt is left behind in the surface water making it saltier and denser so that it settles to the bottom. In the Mediterranean, this settling is at a rate of 2 million tons per second. The denser water flows out along the bottom through the straits of Gibraltar and is added to the eastward flowing waters of

the deep Atlantic. Of course, the surface water of the Mediterranean must be replenished, and the Atlantic supplies most of the water for this purpose, as is evidenced by the fact that water flows in through the surface of the straits at a rate of 2.3 mi/hr.

23-14 Oceans and the Weather

As we have seen, the basic cause of all air flow is a difference in temperature. When the land is warm and the adjacent ocean is cool, as is usually the case on a summer day, a convection current is set up as in Fig. 20-6. The air over the land is heated and rises while that over the cooler water becomes colder and more dense and falls. The cool air then flows inland to take the place of the rising air, and this accounts for the sea breezes that blow inland during the day in many coastal regions. The process is reversed when the water is warm and the land cool. The air over the water is warmed and rises and the flow of air is from the land to the water, which accounts for the land breeze often found at night along coastal regions.

These convection currents are the means by which large quantities of heat are exchanged, and they are partly responsible for the moderate temperatures that are found in coastal areas. While inland average temperatures may vary by as much as 62°F from winter to summer (as in Bismarck, North Dakota—January 8°F, July 70°F), the average temperatures in Seattle, which is near the ocean, at the same latitude vary from 41°F in January to 66°F in July, a difference of 25°F. We have already seen how the climate of the western United States is modified by the moist air coming from the North Pacific. As this air sweeps inland, it must rise across the Cascades or Sierras, which cause it to drop much of its water and in the process to release heat. Along the western shore line of North America, the moderate coastal climate extends only over the relatively narrow belt between the Pacific and the mountains. In Europe, however, there are no coastal mountains, so the coastal weather is carried hundreds of miles inland and thus modifies the climate over great areas of that continent.

23-15 Summary

The sun, 93 million miles away, is the source of practically all the energy presently available here on earth. The body of the sun is composed largely of hydrogen and helium. The surface of the sun is made up of three parts: the 200-mile thick photosphere, which is what we ordinarily see; a 5000-mile thick chromosphere; and the outermost nebulous corona seen only during an eclipse or with special apparatus.

The greenhouse effect depends upon this principle: short-wave radiation can penetrate the atmosphere and be absorbed by the earth, but the long-wave radiation of the earth cannot penetrate the atmosphere and so is trapped. A section of the atmosphere especially enriched with ozone, the ozonosphere, together with water vapor, limits the amount of ultraviolet radiation that penetrates the atmosphere, while water vapor and carbon dioxide limit the escape of long-wave radiation from the earth. These two processes thus establish the heat balance of the atmosphere.

The mass of water vapor per unit volume in the air is the absolute humidity. At any given temperature, the maximum value of the absolute humidity is the saturation humidity, whereas the ratio of the absolute humidity to the saturation humidity expressed in percentage is the relative humidity.

Since warm air can hold more moisture than cold air, whenever air is cooled there is an increase in the relative humidity, possibly to the point where saturation and rain occur. When air that is relatively clear and dry rises or falls, the change in temperature per unit change in elevation is called the dry adiabatic lapse rate. If the air is undergoing condensation of some of its moisture during the change in elevation, the change in temperature per unit change in elevation is known as the moist adiabatic lapse rate. The heat derived by a rising air mass from the condensation of water vapor allows it to return to an equivalent elevation, warmer and drier than before.

The areas from which large air masses originate are known as source regions, and the forward moving edges of these air masses are known as fronts—cold fronts or warm fronts depending upon which type of air they precede. The movement of these air currents across the earth is affected by the rotation of the earth and the resultant Coriolis effect, which deflects winds to the right in the Northern Hemisphere and to the left in the Southern Hemisphere.

The global circulation of the atmosphere that is powered by exchanges of heat energy greatly affects the circulation of the oceans. Because the mass of the oceans is so great and because the specific heat of water is much greater than that of air, the oceans transport the major share of the heat energy that makes large areas of the earth habitable.

EXERCISES

A. KEY TERMS TO REMEMBER

absolute humidity
chromosphere
cold front
Coriolis effect
corona
dew point
dry adiabatic lapse rate
greenhouse effect
insolation
moist adiabatic lapse rate
photosphere
polar front
relative humidity
saturation humidity
source region
warm front

B. QUESTIONS ABOUT HEAT IN NATURAL PROCESSES: I

1. Which of the short-wave processes in Fig. 23-4 cause the sky to appear blue?

2. In Fig. 23-4 show that the amount of energy radiated from the planet earth is the same as that which the earth receives.

3. At the surface temperatures existing on some of the outer planets, it is possible that ammonia (NH_3) or methane (CH_4) are in the liquid state. Outline what you think would be the difficulties of having a system of life based on either of these chemicals rather than on water, as it is on the earth.

4. In Fig. 23-10 it appears that the outgoing terrestrial radiation exceeds the incoming solar radiation. Is this actually true? If your answer is "no," how do you explain the graph in the light of such an apparent contradiction?

5. Suppose that you were in charge of an interplanetary rocket launching project. If the rocket were fired from your locality vertically upward, what eastward velocity component would it have?

6. Predict the rainy areas of the world by looking at a map that shows little more than where mountains exist.

7. In discussing Fig. 23-6, we said that the temperature on the "back side" of the mountain was warmer because it was drier. Should we simply have said it was warmer and drier? Explain.

8. How can the relative humidity be only 90 per cent when it is raining?

9. Suggest several regions where descending air is commonplace and give the reasons why the air is descending.

10. Why is the moist adiabatic lapse rate less than the dry adiabatic lapse rate? Although this was not discussed, why would you expect this lapse rate to be variable?

C. PROBLEMS ABOUT HEAT IN NATURAL PROCESSES: I

1. What is the relative humidity of air at a temperature of 25°C if moisture just condenses on a glass that contains water at a temperature of 10°C?

2. If air has a temperature of 25°C and a relative humidity of 62 per cent, approximately how much water will condense per cubic meter if the air temperature is reduced to 5°C?

3. If a dry-bulb thermometer reads 71°F and the relative humidity is 45 per cent, what will be the reading of a wet-bulb thermometer? Ans. 58°F.

4. Air at sea level at a temperature of 73°F has a dew point of 55°F. Suppose this air moves up the side of a mountain range to an elevation of 8000 feet and then drops down the back side to sea level again and that the dry adiabatic lapse rate is 5½ °F/1000 ft, the moist adiabatic lapse rate is 2½ °F/1000 ft, and the dew point decreases 1°F/1000 ft. Determine:
   a. The elevation of the base of the clouds formed as the air rises.

b. The air temperature and the dew point at 1000-foot intervals up and down the mountain.
c. The elevation to which the air would have to be raised in order to produce snow.
d. Approximately how much moisture is condensed per cubic meter from the air mass by the time it reaches the crest of the mountains.
e. The elevation at which the air has the lowest relative humidity.
f. The elevation or elevations at which the air has the highest absolute humidity.

5. In Fig. 23-6, to what elevation would the air have to be raised in order to produce snow? Ans. 9333 ft.

6. Determine the average annual rainfall for your location. From this determine the approximate amount of heat in Btu's that are delivered each year to the air over each square foot of the earth at this point.

7. Explain what would occur in your backyard if the air temperature were 45°F, the dew point 40°F, and the ground temperature 38°F. Assume the sky is clear and there is little or no wind.

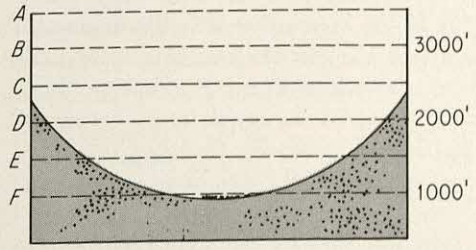
8. Explain what would occur in problem 7 in the following case: air temperature, 34°F; dew point, 31°F; ground temperature, 29°F; clear air, no wind.

9. Figure 23-20 represents a valley with changeable air temperature and humidity conditions. Explain what would happen in each of the following cases:

a. A—46°F air temp.
B—44°F
C—42°F
D—40°F
E—38°F
F—36°F
dew point 40°F

b. A—34°F
B—36°F
C—38°F
D—40°F
E—42°F
F—44°F
dew point 38°F

Fig. 23-20. A valley with variable conditions.



Twenty-Four

Heat in Natural Processes: II

Any of us who have seen hot springs or geysers or the evidences of volcanic activity must have guessed that the center of the earth is very hot. The fact is, however, that we know very little about the center of the earth—what its pressure is, its temperature, or exactly of what it is made.

Explorations into the earth by mining and drilling operations have given us some widely differing figures for the temperature change with depth inside the earth. In one of the deep oil wells in California the temperature at a depth of 16,000 feet is 400°F. In some places the increase in temperature is as high as 1°F in 30 feet, but in South Africa it is 1°F in 250 feet. So while there are great differences in the temperature gradient in various localities, a statistical evaluation of all the data leads to the conclusion that the average temperature gradient is about 1°F per 60 feet.

The fact that we know the temperature gradient in the earth, within some limits, gives us an opportunity to estimate the temperature at the earth's center. In doing this we must extrapolate. If we assume that the rate of temperature change remains constant throughout the earth, we find that this leads us to the seemingly impossible temperature of 350,000°F. When we think of the temperatures of the sun and recall that the "flaming" surface of the photosphere is only 6000°C or 10,000°F, it seems certain that our extrapolation is invalid. We should have some reservations about this procedure anyway, for we have seen that the temperature gradient at different parts of the earth varies widely. Since this is so, it is all the more unlikely that the temperature gradient remains constant throughout the great distances and variety of materials to be found between the surface

and the center of the earth. It is probable, too, that as the pressure on the rocks changes with depth, the heat-conducting properties of the rocks also change. And the fact that the interior of the earth probably contains a higher percentage of metallic elements than the surface, also might change the heat-conducting properties of the earth's center. Since we have direct evidence of the earth's temperature gradient in only a few places and to a maximum depth of only 1/1300 of the earth's radius, extrapolation appears to be a poor method to determine what the temperature is at the center of the earth.

The more trustworthy evidence from earthquakes (that is, seismological evidence) indicates that the rocks behave pretty much as a solid to a depth of 1800 miles. If this is the case, the temperature at this depth must be less than the melting point of the rocks under the pressures existing there. This, of course, poses the new problem of determining the temperatures at which rocks will melt under these enormous pressures. We shall have to learn a great deal more about the earth and the material of which it is composed before we can answer with certainty the question, "What is the temperature at the center of the earth?" It may seem strange that we know so little about the interior of the earth upon which we spend our lives, but we do not have to go far in science before we find out that the simple questions do not necessarily have simple or easily found answers. The best answer we can get at the moment from the experts is their estimate that the temperature at the center of the earth is between 2000°C and 5000°C.

24-1 Sources of the Earth's Heat

When we try to determine the source of the earth's heat we also face some difficult problems. The old theory was that the earth is in the process of cooling down after its formation from the sun. This theory supposes that the earth was originally at the sun's temperature but has since cooled and formed a hard crust around a still intensely hot core. Just how much heat is being conducted away from the core? When we measure the heat conductivity of the rocks of the earth and use an average temperature gradient, we find that the average heat flow is 1.2×10^{-6} cal/cm^2/sec. Although this is not large, it is nonetheless much more than all the heat brought to the surface by the earth's volcanoes. This amount is too small to have much effect on surface temperatures, however, since solar radiation supplies about 3500 times as much heat.

The discovery of radioactive materials—uranium, thorium, and radium—which give off heat continuously, has complicated the picture and introduced a new theory about the source of the earth's heat. Granite contains many different elements and among these are three that contribute to the

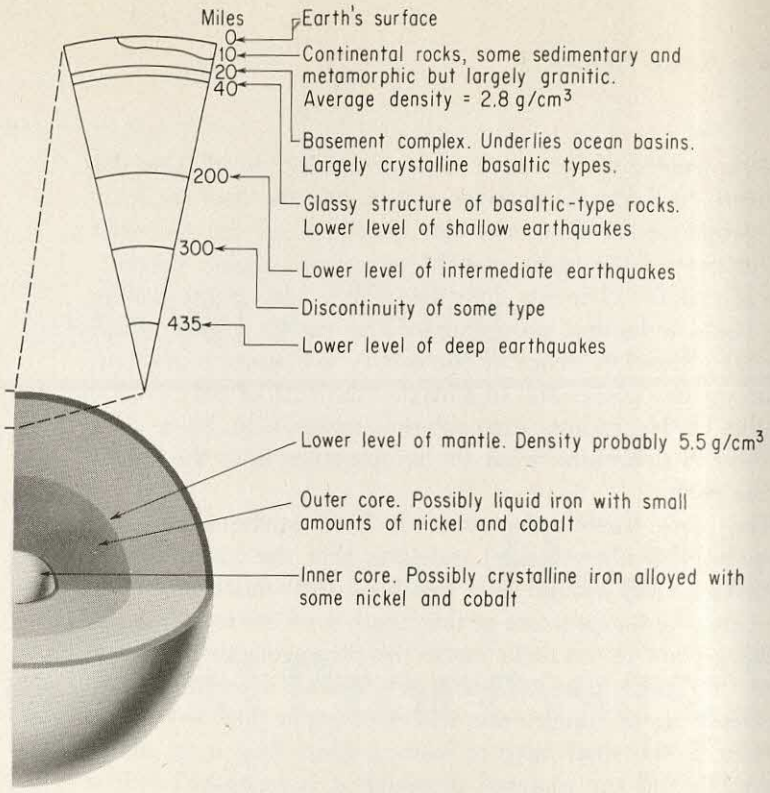


Fig. 24-1. Cross-section of the earth. The mantle is approximately 1,780 miles thick, is solid, and is composed of silicate material, possibly $(Mg,Fe)_2SiO_4$ plus some FeS. Its mean density is 4.5 g/cm³ and it contains 67.8 per cent of the earth's mass. The core is approximately 2,150 miles thick, has an average density of 10.7 g/cm³, and contains 31.5 per cent of the earth's mass.

heating of the rock through radioactivity. There is 9×10^{-6} g of uranium in a gram of granite and also small quantities of thorium and a radioactive isotope of potassium. It has been computed that the uranium in a 19-mile thickness of granite, by its natural radioactive decay, would supply an amount of heat equal to that flowing out of the earth, and this could certainly explain the heat of the earth, for this thickness of granite undoubtedly exists. However, the earth is 3958 miles in radius and if the radioactivity in granite is the source of the earth's heat, we will have to assume that the radioactive materials are not distributed uniformly but are concentrated largely in surface rocks. If they were distributed uniformly there would be much more radioactivity available for heating and much more heat given off than we actually find coming from our earth. Thus, if radioactivity is the source of the earth's heat, it raises new problems that are as yet unsolved.

While radioactivity may be one answer to our question regarding the source of the earth's heat, another proposal has been set forth based upon the high pressures that exist at the earth's core. The familiar fact that gas is heated as it is compressed led early scientists to believe that the cooling and shrinking process undergone by the earth increased the pressure, and therefore the temperature, of the rocks at the center of the earth. If the earth is shrinking in size or in some other way increasing the pressure on its core, this may at least partially explain the source of the earth's heat. The definite cause of the earth's heat, however, will have to remain, for the present, one of the great unanswered questions of science, and perhaps the theories of the origin of the earth, which we will discuss in later chapters, will give

us the closest thing we have to an answer. When the answer is finally discovered, it will probably be found to be a combination of theories that have been contributed by scientists in many different fields of science.

24-2 Magma and Molten Rock

Magma is the name given to molten rock within the earth. When it appears on the surface we call it *lava*. In the early days of scientific investigation, it was thought that magma rose through tubes or zones of weakness from the molten interior of the earth. As we have seen, however, earthquake studies indicate that the earth reacts essentially as a solid to a depth of 1800 miles. It is hard to conceive of molten rock having to move so far in order to produce a lava flow or a volcano. Lava, therefore, must originate as magma not too far below the earth's surface. Basalt rocks that melt at the surface at temperatures of 1250°C will melt at a temperature of approximately 1400°C at the higher pressures existing at a depth of 22 miles, and this is approximately the temperature of lava as we find it. Since rocks at a depth of 22 miles still act as solids to earthquake shock waves, it is possible that temperatures at that depth approach, but do not exceed, the melting point of rock. Only a little bit of heat needs to be added to the rocks, then, for them to melt and become fluid magma.

In the melting of rock the same principles apply as in the cases considered in Chapter 22; that is, rock expands when heated and during melting, and an increased pressure makes it more difficult to melt and increases the melting temperature. An additional complication arises because melting rock is a mixture of many materials and will not behave as a simple compound. It is known, for example, that granite has a melting point of 1100°C when it is completely dry, but when it is a mixture of 6 per cent water and 94 per cent granite, its melting point is 760°C. We can deduce from this that magma may have quite different melting points depending upon its composition. Possibly then, the movement of various gases and fluids throughout the interior of the earth can change the melting point and in itself set off the cycle that results in volcanic activity. The earthquake activity that often precedes the activity of a volcano may provide channels by which water, gases, and other fluids may reach rocks near the melting point and, in combining with them, change their melting point so that liquid magma results. The final explanation of the process by which magma is generated is unknown and forms still another great unanswered question of science.

Quartz is one of the important minerals in granite and will serve to illustrate the melting process as it applies to any rock. We must remember that this is a single mineral as contrasted with magma, which is a mixture of many min-

erals, liquids, and gases. Quartz has a melting point of 1470°C, and this temperature must be close to that which we would have to consider in the case of magma. With such high temperatures, the specific heat of quartz is not exactly the same as at ordinary temperatures. For instance, at 1400°C the specific heat is 0.294 whereas at 100°C it is 0.237. Because of this difference in the specific heat, any graph that shows the total number of calories of heat contained in the system compared with the temperature will not be a straight line relationship, as we can observe in Fig. 24-2.

We have already learned that to change a solid to a liquid we must add the heat of fusion to the material. Generally, rocks, including quartz, expand during the melting process, and this expansion, under pressure, takes additional heat. We should not be surprised, then, to find that the heat of fusion also varies with pressure. If we plot volume against temperature for the melting of quartz, we get the curve shown in Fig. 24-3. In order to change the solid quartz to liquid quartz, we have to add 59 calories per gram for the heat of fusion. Any additional heat added after the quartz is melted raises the temperature of the resulting liquid and causes it to expand, but before much additional heat is added to the liquid, convection currents will probably be set up within the liquid that will be effective in transporting the excess heat to other solid quartz or to the surrounding rock, causing it to melt. As a result, the extruded rock that comes from volcanoes and fissures should all be near the melting point and this prediction seems to be borne out by observations around the earth.

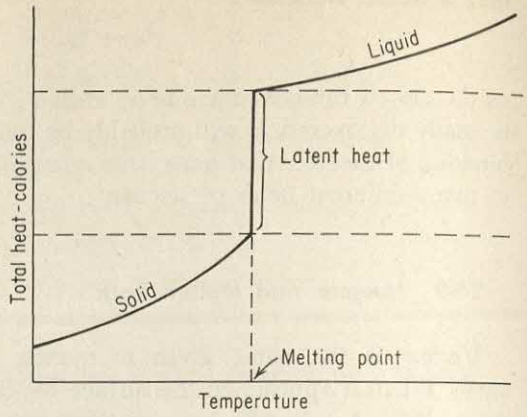


Fig. 24-2. Heat-temperature relationships for melting quartz.

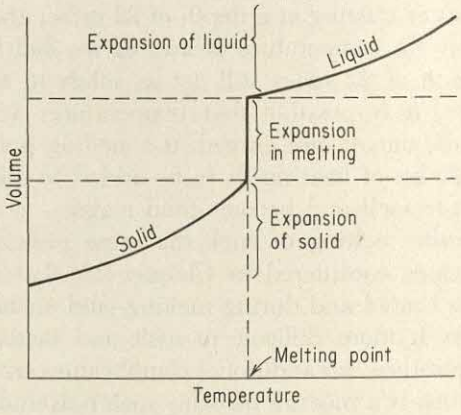


Fig. 24-3. Volume-temperature relationships in the melting of quartz.

24-3 Intrusive Magmatic Activity

Most of the movement of magma is within the earth and never seen by man. This movement in which the molten rock forces its way into and through cracks in other rock structures is called *intrusive activity,* and the rock structures that result are called *intrusive formations.* Only when erosion exposes these formations or man digs into the earth far enough does he find the results of such movement. We will not attempt to name all the different types of intrusive formations, but there are some that we should know because they are so readily recognized and so common (Fig. 24-4). When magma has forced its way in between rocks that have already been formed into some type of layered structure, it is called a *sill.* A sill may be less than an inch or more than 500 feet thick, but it is not considered a sill unless it is more than ten times longer than it is wide and high. By the way it is formed, we can see that a sill is made up of rock that is younger than the adjacent rock.

Sometimes a large magmatic dome pushes up the strata into which it has intruded, and if such a formation cannot be called a sill because of its lack of extent, it is a *laccolith.*

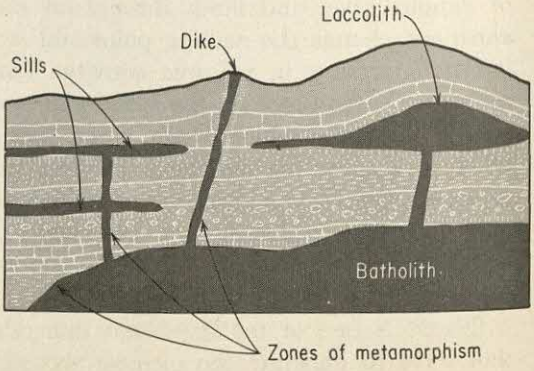


Fig. 24-4. Intrusive geological formations.

Fig. 24-5. Photograph of Shiprock, New Mexico, showing dikes radiating from the volcanic neck. (Fairchild Aerial Surveys, Inc., L.A.)

This type of structure may be fed from its own volcano tube or may be a sort of blister on the top of a sill.

If an intrusive mass cuts across the prevailing rock structure or fills a fissure in pre-existing rocks, it is called a *dike*. Dikes are much the same as sills, except that they intrude right through the existing rock rather than parallel to it. Dikes come in various sizes, ranging in thickness from a fraction of an inch to hundreds of feet, while their length may be as much as 65 miles or more (Fig. 24-5).

When magma moves through other rocks, the heat, the accompanying gases, and the chemically active fluids that accompany it act to change the neighboring rocks by both physical and chemical means into different rock structuring and sometimes into a material known as *metamorphic rock*. Marble is an example of such a rock that is produced by the action of great pressures and high temperatures on limestone. Intrusive structures have associated zones of metamorphism that sometimes contain concentrations of certain minerals that have been removed from the parent rock by the metamorphic agents and have been formed into the mineralized veins sought by prospectors and mining geologists.

In the Sierra Nevada mountains there are very large areas of granite, an intrusive rock that in this instance has no apparent lower floor or surface. Such rock is called a *batholith*. We can distinguish between a laccolith and batholith by noting that the former has a definite floor. Batholiths are often found in mountain ranges. The largest known batholith covers 100,000 square miles, forms the backbone of the coast range of British Columbia, and extends some 1250 miles northward into Alaska. These large masses of rock are usually composed of granite, and this fact is thought by many geologists to be evidence that they were formed deep within the earth's crust. Batholiths are usually located where there is strong folding of the mountains, but this does not mean that the batholith produced the folding. As a matter of fact, it is equally possible that the folding of the moun-

tains so weakened the supporting rocks by fracture that they were quite readily dissolved or melted away by the intruding igneous mass. There are still many unsolved questions concerning the formation and structure of a batholith, including the one of just how the granite itself was formed. Geologists have watched many igneous rocks form, but they always result in basaltic, not granitic, types of rock. Nevertheless, the evidence available at present indicates that granite is also a true igneous material. The processes involved in the formation of granite will remain in doubt, however, until further research can help us find the correct answer.

Why magmatic rocks tend to rise in the earth's crust is also an unsolved question. At present there are two theories to explain this rise. Some scientists compare the rise of magma to the squeezing out of toothpaste from a broken tube that is being compressed. They theorize that as the earth cools it gradually shrinks and that this shrinkage takes place from the outside inward, resulting in increasing pressure on the interior rocks. The molten magma, in seeking to adjust to this increasing pressure, then squeezes out wherever it finds a zone of weakness in the earth's crust.

The second explanation could be called the "coke bottle" hypothesis. It assumes that the magma itself is charged with gases—predominately water vapor—which greatly expand when the pressure is released. When a fissure is opened by earthquake activity or by the melting of the rock through contact with the magma, and the magma can find room to expand, the gases keep forcing the magma outward, and eventually it may reach the surface and cause a volcanic eruption accompanied by the explosive escape of water vapor, carbon dioxide, and the other volcanic gases.

24-4 Chronology of Geologic Events

Let us pause in our consideration of the earth's heat to summarize some information and ideas we have learned about rock structures and their formation. Two of the principles involved are so logical that you may have discovered them *for yourself*. The first is the *principle of uniformitarianism*. This states that geologic events of the past took place in much the same way and by the same processes that are at work on the earth today. Just as we saw in Chapter 4 that conglomerate is formed close to a shore, whereas sandstone is formed somewhat further out (and presumably at a greater depth), we can assume that this same thing has held true throughout the earth's history.

A second principle is the *principle of superposition*, which states that a rock stratum on top of a second stratum is younger than the deeper rock, provided that earth movements have not disturbed the beds.

We can combine these two principles to give us some idea

about many ancient geologic happenings. Suppose we find sandstone beds resting on top of conglomerate. A geologist might reason as follows. Since conglomerate is underneath the other sedimentary rocks, it is the older of the two. Since conglomerate was formed in shallower waters (or closer to a shoreline) than sandstone, the earth's surface here had first been close to a shore and later farther from the shore (first in shallow water, then in deeper water). If we take the level of the oceans as our reference level, we can say that the land (or ocean floor) was sinking during the formation of these rocks.

Other ideas are summarized by the rock cycle in Fig. 24-6. The outer circle represents one path of events that

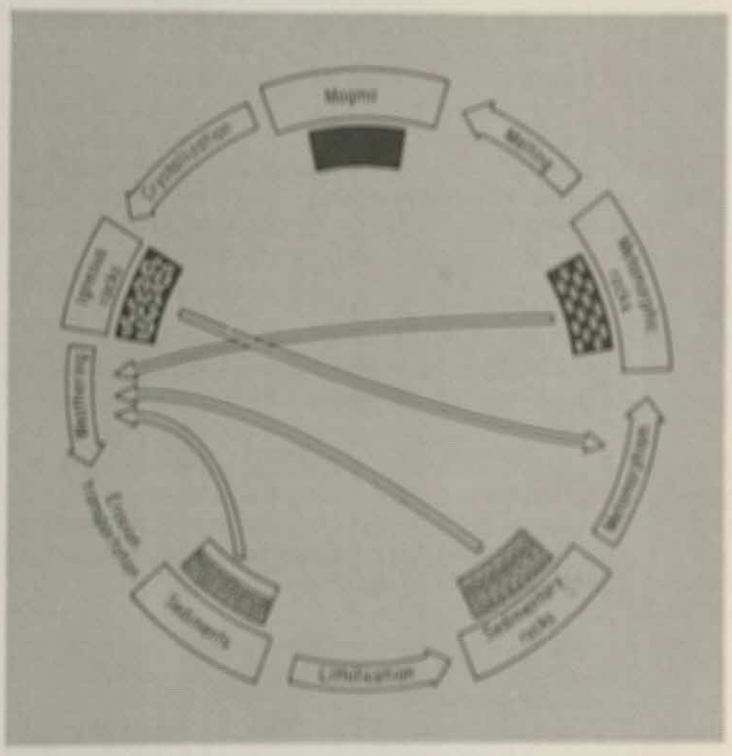


Fig. 24-6. The rock cycle, shown diagrammatically.

might occur. Magma when cooled and hardened crystallizes to igneous rocks, which weather to produce sediments. These in turn may be cemented together (lithified) to become sedimentary rocks. The sedimentary rocks in turn may be metamorphosed and the resulting metamorphic rocks may finally become magma again. The cycle may be interrupted at various points along its path and may follow the course indicated by one of the arrows shown in the interior of the cycle.

Now you are in a position to put all of these ideas together to see how a geologist would deduce the order of

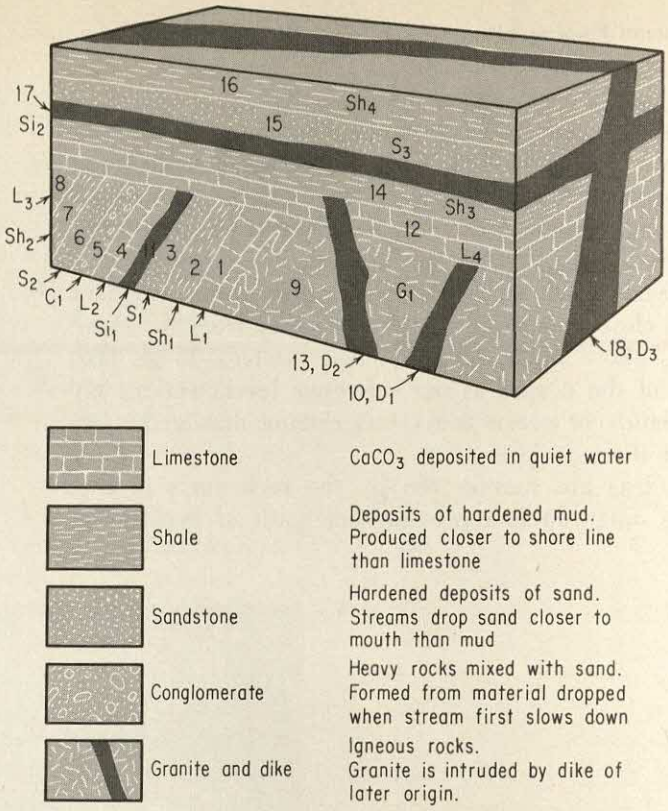


Fig. 24-7. Determining the order of geological events.

geologic events. Careful study of Fig. 24-7, which provides a hypothetical cross section of a portion of the earth's crust, will give you a much better understanding of geologic reasoning. The numbers on the figure suggest the probable order of occurrence.

The order of geological events in Fig. 24-7 may be deduced as follows:

1. Deposition of limestone L_1 in a deep sea. L_1 came before granite or Sh_1 because it is intruded by the granite and is underneath the layer of shale.
2. Deposition of shale Sh_1. Since shale is deposited nearer the shore, the sea must have become shallower, caused by the land rising.
3. Deposition of sandstone S_1. Deposited still nearer shore, land still rising.
4. Deposition of limestone L_2. Zones of metamorphism on either side of sill Si_1 would indicate intrusion rather than a surface flow of Si_1. Land must have again fallen.

5–8. Deposition of C_1, S_2, Sh_2, and L_3, in this order.

9. Intrusion of granite G_1 as a batholith, which occurred some time after L_1 was deposited, but was probably a deep intrusion that took place after steps 5–8.

10–11. Intrusion of dike D_1 and sill Si_1. The basalt of the dike is younger than the granite. Tilting of beds and granite indicates possible mountain building

to the right, which was followed by weathering of the land. Land worn down to a relatively smooth surface called an unconformity. Hard and soft rocks are worn down to same elevation indicating a long period of erosion with the land at the lowest elevation of a water surface, probably sea level.

12. Deposition of limestone L_4.
13. Intrusion of dike D_2.
14–16. Deposition of Sh_3, S_3, and Sh_4.
17. Intrusion of sill Si_2.
18. Intrusion of dike D_3.

24-5 Types of Volcanic or Extrusive Activity

If magma in its movement is forced above the surface of the earth, it is said to be *extruded*, and it makes its presence known in a number of ways. The most spectacular of these is probably the *pyroclastic* or *cinder cone*. These cones typically consist of cinders and fragments of rock that are thrown upward and solidified while still in the air. The cones are relatively small (seldom exceeding 1000 feet in height), and their sides are sloped as steeply as possible without sliding. There are many examples of this type of volcanic activity in our western states. In the San Francisco mountains, just south of the Grand Canyon of the Colorado, there are over 200 of these cinder cones (Fig. 24-8). The Mexican volcano, Parícutin, is a recent example of this type of magmatic activity. The formation of this cinder cone dramatically took place in the presence of Dionisio Pulido, who at noon on February 20, 1943, while plowing his cornfield noticed a thin wisp of smoke rising from one of the furrows. By four

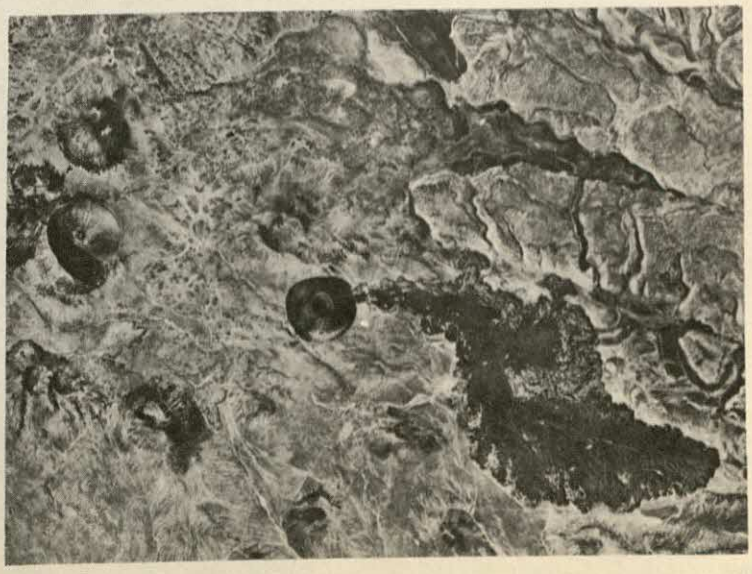


Fig. 24-8. An aerial view of cinder cones north of Flagstaff, Arizona. (Fairchild Aerial Surveys, Inc., Los Angeles)

o'clock that same afternoon dense clouds of smoke were rising, and explosions occurred every few seconds. Five days later the cone was 300 feet high and after a year it had grown to 1400 feet. Two days after the start of the eruption, lava flowed from a fissure about 1000 feet north of the cone, and fifty days later this flow had advanced about a mile. The activity ceased after nine years and Parícutin became just one of the many dead cinder cones that are found in the vicinity. Geologists believe that the eruption of this volcano was typical of many that have occurred in the geologically recent past in the western United States.

Another type of volcano cone results from the quiet extrusion of lava. When lava flows are piled one on top of another around a central vent, they form a *shield volcano.* Mauna Loa, one of the largest of these structures, rises from the floor of the Pacific nearly 6 miles to its crest, and its lavas cover some 2000 square miles at sea level and form more than 50 per cent of the island of Hawaii. The slope of the sides of a volcano is governed by the viscosity of the extruded lava. If the lava is very fluid, the volcano will have a very gentle slope, whereas if the lava is thick and viscous, the volcano will have steeper sides. Since the viscosity of the lava depends upon its chemical composition and its temperature, the slope and profile of a shield volcano are the result of a complex series of circumstances. These volcanoes usually form gently rising domes with slopes ranging from 2 to 10 degrees.

Very often these two types of activity are combined. A volcanic vent may periodically erupt, spreading cinders, ash, and other pyroclastic material over its sides, and later, or sometimes even associated with this eruption, lava may flow out of fissures along the sides of the volcano. This combination of explosive ejection and quiet flow produces what is known as a *composite cone,* and these cones form some of our most beautiful mountains. Mt. Fujiyama in Japan, the volcano Mayon in the Philippines, Mt. Rainier (Fig. 24-9),

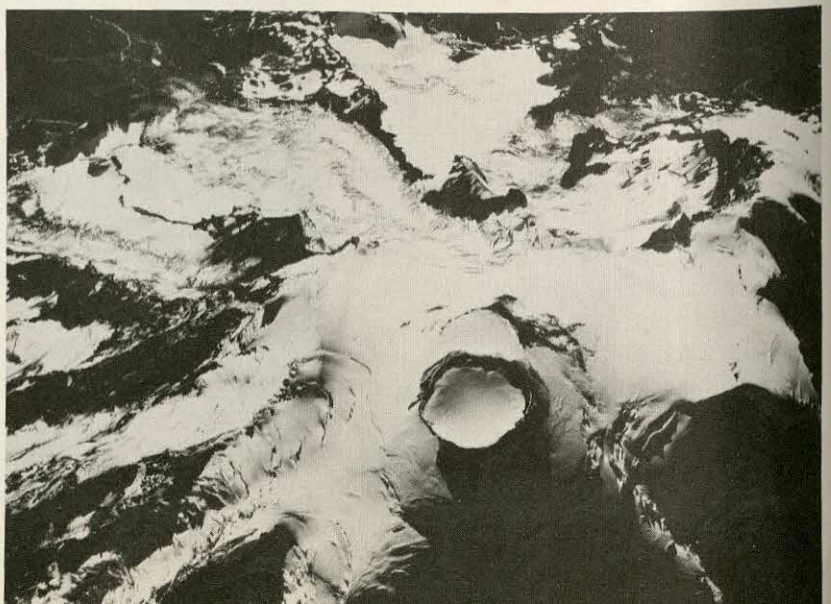


Fig. 24-9. The crater of Mt. Rainier, Washington. (Boeing Airplane Co.)

Mt. Adams, Mt. Hood, Mt. Shasta, and others of our own Cascade Mountains are examples of composite volcanoes.

Sometimes a very fluid type of lava, called a *fissure flow*, pours out from cracks in the earth's surface and flows like a river, following valleys, filling depressions, and leveling the contours of the countryside. Nearly the entire eastern half of Oregon and Washington, much of Idaho, and parts of Nevada and California are covered by such a flow. Although these flows are geologically important and cover hundreds of thousands of square miles of the earth's surface, only one authenticated fissure flow has occurred within historical times. On June 11, 1783, there were immense outpourings of lava along a 10-mile fissure in Iceland. This flow, which had been preceded by a series of earthquakes centering near Mt. Skaptar, poured into the valley of the Skapta River, dried it up, and dammed its channel (which had been from 400 to 600 feet deep). Within a week a second flow took place, and on August 3 of that year a third flow occurred that filled a lake and oozed across the countryside in tongues up to 15 miles wide and 100 feet deep. Some flows extended more than 25 miles before the lava cooled and began to choke off the fissure, and as it began to clog up, 22 small cones formed along its length. Spewing forth gases and globs of molten lava, these cones relieved the last of the pressure, and the activity died away.

Occasionally a volcano that has already formed a sizable mountain will undergo a violent eruption that leaves a great circular pit known as a *caldera*. Some calderas are formed by gigantic eruptions that literally blow the top off the mountain, although most explosion calderas seldom exceed one mile in diameter. Other calderas, such as that at Kilauea in the Hawaiian Islands, result from a subsidence of the level of the magma beneath the crater or along a circular fracture. Some larger calderas, however, have a more complex origin. Crater Lake in southern Oregon is thought to have resulted from both a gigantic explosion and a subsidence of the supporting magma. The crater, which, of course, is not the eruption crater of a volcano at all, is over 5 miles in diameter and some 4000 feet deep. The lake in the caldera is 2000 feet deep, and the fractured walls rise another 2000 feet above the lake surface. From the probable shape of the original Mt. Mazama (which is the name given to the pre-caldera mountain) and the present outline of the lake's rim, it is estimated that some 17 cubic miles of material were blown out or drained away when the caldera was formed. An island now jutting out of the lake is obviously a cinder cone of geologically recent origin, and a lava flow has been found on the bottom of the lake, indicating that the mountain was active after the major eruption and subsidence.

It seems certain that the behavior of magmas must be related, among other things, to their chemical compositions. Table 24-1 summarizes the composition of two different lavas, known as "acid" and "basic," although this is not di-

rectly related to pH in the sense the term was used in Chapter 15. Lavas are considered "acid" if they have more than 60 per cent of SiO_2 and "basic" if they have less than this amount of SiO_2 in the free state or in combination. No simple correlation exists between these percentages and the properties of the emerging lava, since the various minerals that make up the magma will have differing melting points, and crystallization of some minerals will occur in the magma chamber or the conduit. As a result, the character of the eruption and any resulting lava will not be determined by the bulk chemistry, but almost exclusively by the composition of the emerging portion of the lava.

TABLE 24-1 COMPOSITION OF LAVAS

	Typical "Acid" Lavas, %	*Typical "Basic" Lavas, %*
SiO_2	70+	49
Al_2O_3	12–15	16
Na_2O and K_2O	6– 8	Less than 3
MgO	Less than 3	6
CaO		9
Fe_2O_3 Fe_3O_4		12

The viscosity of the lava is one of the major factors controlling the character of the eruption. In cases of a low viscosity, gases can readily escape from the magmatic material so that little explosive material is produced. There is more explosive loss of the volatile materials as the magma becomes more viscous. This process is accompanied by an increase in the amount of fragmentary material and in the violence of the explosions. Table 24-2 summarizes these changes in properties.

TABLE 24-2 PROPERTIES OF LAVAS

	"Acid"	*"Basic"*
Viscosity	Greater, "stiff"	Less, "fluid"
Rate of flow	Slow	More rapid
Liberation of gas	With difficulty	Easily
Type of eruption	Explosive	Quiet
Resulting rocks	Rhyolite	Basalt
	Light grey or pink	Dark colored
	Intermediate Lavas	
	Andesite (from Andes Mts.)	

Besides viscosity, several other factors exert an influence upon the structure and form of the volcano. The form of the magma feeding conduit is perhaps the most important of these. The shape of the resulting structures are profoundly affected by whether it is a fissure a mile or more long or a tubelike conduit. The length of time involved and the num-

ber of eruptions taking place at the same vent or from the same magma source are also important.

24-6 A Cycle of Volcanic Activity

Suppose we trace the events that usually precede and follow volcanic eruptions. The first sign of an impending eruption ordinarily is the gradual rising of the floor of the volcanic crater. Simultaneously there may be an increasing number of small localized earthquakes. An actual swelling of the mountain may occur so that instruments can detect a slight tilting away from the crater area. Under increasing pressure, magma forces its way higher and higher, and temperatures inside the crater will show that the material is becoming hotter and more fluid. As the magma approaches the surface, it gradually absorbs ground water and possibly oxygen, which lower the melting point of the magma and cause it to become superheated and charged with gases. The high temperature, coupled with the easy transportation of the heat by convection currents in the fluid magma, melts the covering rock or dissolves it away, and the magma breaks out through the surface of the crater or through a fissure in the volcano's side. The lava that spills out will reduce the pressure on the remaining magma, and the gases it holds tend to expand and blow off their excess pressure in a rush that is something like the "whoosh" of gas that ensues when we shake up a bottle of carbonated water and then take off the cap. The blow-off of gas will throw out of the crater any debris that was left by prevous eruptions or by erosion (Fig. 24-10).

As the pressure is released and the magma starts to cool down, material tends to fall back into the crater and clog it up, and before the volcano returns to a quiet state, it will probably periodically throw out some of these cooled materials in a dark ash-type eruption. Gradually the volcano settles down and becomes quiet until it can again accumulate

Fig. 24-10. Aerial view taken March 29, 1947, of Mt. Hekla, Iceland, as the volcano sends huge clouds of smoke and tons of lava and rock skyward. (Wide World Photos)

enough heat to start the cycle all over again. Just as there are all sorts of people so there are all sorts of volcanic mountains, and the behavior of many of them varies from that we have described. Nevertheless, the activity of volcanoes depends upon variables that we can largely measure, and there is hope that as we learn more about these fascinating mountains we may be able to predict their eruptions and possibly even control them.

24-7 Types of Ejected Magmatic Material

Many types of material are ejected from volcanoes. Dust and ashes, of course, are thrown upward, sometimes to a great height, but we should be certain when we speak of volcanic ash that we are not referring to the products of true combustion, for there is little real burning associated with the crater of a volcano. During the eruption of Krakatoa in 1883, when four cubic miles of material were blown skyward, the dust rose to a height of 17 miles, so high that it spread completely around the earth. It is known to have affected the sunset coloring and to have reduced the amount of solar radiation arriving at the earth for a period of more than three years. As early as 1783 people noted a connection between volcanic eruptions and the weather. During that year the volcanoes Asama in Japan and Laki in Iceland erupted with violence, and the immense amounts of volcanic dust they shot into the upper atmosphere apparently reflected so much of the sun's radiation that the following winter was one of the most severe on record in North America and Europe. The following summer, average temperatures dropped further than they did during the next 150 years. One of the country's early scientists, Benjamin Franklin, published a description of some of the phenomena that resulted from these eruptions and speculated on the possible connections between volcanic eruptions and the weather.

Particles too large to be called volcanic dust are called *lapilli*, and still larger bits of matter are called volcanic *bombs*. Bombs are streamlined rocks that are blown into the air while molten and then cool and harden while under the influence of air friction. When we see a motion picture of a volcano in eruption, we can often see the volcanic bombs being formed. As the material is thrown upward, its color shows that it is very hot and probably fluid, and as it traverses its path through the air, we can watch its color change from a bright red or orange to darker shades and finally to black, which indicates that the erupted material has hardened. The change in color with temperature that we witness here is another illustration of the information incorporated in Fig. 12-6.

The bombs and the fragmented material usually fall near the volcanic vent and tend to build up a cone. The lighter

materials, more affected by the wind, are blown farther from their source. Dust showers from volcanoes may be heavy enough to be very destructive to plant and animal life, but after this dust falls to the ground, weathering may change it into a very rich soil. Farms in volcanic regions extend right up the sides of volcanoes to take advantage of this fertile volcanic earth.

In the year 1902, Mt. Pelée on the island of Martinique in the West Indies erupted with explosive violence. Pent-up gases pushing against an old plug of hardened lava finally broke out sideways, blowing a great cloud of hot gas and molten lava down upon the city of St. Pierre. With the sudden release of pressure, the gas within each fragment of lava expanded and escaped, forcing the cloud of debris, which was heavier than air, to roll downward at a speed of 60 miles per hour. Each of the fragments of rock was surrounded by the very hot gases that were escaping from its surface, and the whole mass descended abruptly upon the city, killing 40,000 people. The type of debris that resulted from this gas-lubricated cloud has been recognized in many other localities, including our own Crater Lake. Since the unsorted nature of these deposits had been very difficult to explain, the eruption of Pelée gave geologists a chance to study their formation and find a satisfactory explanation of their origin. Such a cloud is now known as a *peléan cloud*, and the resulting unsorted material is called *nuée ardente* ("glowing cloud").

24-8 The Earth's Heat and Human Affairs

Volcanoes are not solely destructive. As we have learned, the soil that results from volcanic eruptions is apt to be very fertile and productive. In Latin America the most fertile land is that created by volcanoes, and in Europe vineyards grow high on the slopes of Vesuvius and Mt. Etna.

In Italy, in northern California, and in Java the heat of the earth is being used to produce power. Man has for ages collected the chemicals that are deposited along the sides of gaseous vents, or *fumaroles*, near volcanoes, and not so long ago there was active mining of sulfur in the crater of Mt. Adams in the Cascade Range. The Italians for years have been utilizing the boric acid and carbon dioxide that are brought up in certain of the steam vents in their country, and at present they have a number of power plants converting the energy of the escaping steam into electrical energy. As man uses increasing amounts of energy, he may have to turn more and more to some such source to supply his vast energy needs.

Men's beliefs are often strongly influenced by the environment in which they live. The peoples around the Mediterranean Sea, who were certain that the inside of the earth was molten and very hot, because of the volcanic activity

in their area considered Hell to be an oven of fire and brimstone. Men living in the Arctic, on the other hand, feared that their after-death punishment for evil doings would be exposure to the eternal cold.

The eruption of Vesuvius in A.D. 79 has been of immense help to archaeologists in reconstructing the life of Roman times. The eruption seems to have been abrupt and similar in many respects to that of Pelée. Ash in large quantities quickly buried the towns of Pompeii and Herculaneum, covering them with a protective blanket of dust that preserved them almost intact until modern archaeologists began to unearth these cities.

Our natural fear of volcanic activity is largely justified, for such activity has adversely affected the lives of millions of people. In the eruption of Krakatoa in 1883, the explosion blasted a crater over 1000 feet deep and sent a huge tidal wave across the south Pacific that smashed far up onto the adjacent islands and killed a recorded 35,000 people in the villages located near the seashore. The explosion was heard 2000 miles away, and ships over 1600 miles from the volcano were covered by large quantities of dust three days after the eruption. If one of our Cascade volcanoes erupted with a similar force, the sound of the explosion would be heard as far east as Chicago, and if one of the old volcanoes near Santa Fe, New Mexico, exploded with a similar force, the sound would be heard as far away as New York City. The fissure flow in Iceland in 1783 killed over 9000 people and 230,000 cattle either directly under the lava streams or indirectly by gas poisoning and the famine that resulted from the destroyed crops and the poor fishing. And so we can see that the earth's heat through volcanic activity works both for and against man—providing heat, energy, chemicals, and an increase in soil fertility, as well as death, destruction, and the disruption of the fine heat balance maintained by the atmosphere.

24-9 Phenomena in Lava Flows

We have seen that a flow of lava often accompanies a volcanic eruption and that it may also have its origin in fissures. Lava flows downhill just as any river. Sometimes it will dip into a stream of water or a lake and be cooled so quickly that there is no chance for mineral crystals to form, and a noncrystalline, glassy structure results. This type of rock is known as *obsidian*. When the lava is highly charged with gas and hardens while the gas is in the process of expanding, the resultant rock, called *pumice*, is full of gas bubbles, is frothy in texture, has a low density and may even float on water.

One of the interesting things that often occurs in a lava flow is the formation of a *lava tube*. As the flow proceeds

down a valley, the sides and bottom are in contact with the earth and cool and solidify more quickly than the rest of the flow. The surface also cools rather rapidly for it can lose heat readily by radiation and convection, leaving the inside still molten and able to flow. When the stream of lava is finally stopped at the source, the flow at the far downhill end may still continue drawing lava between the solid walls that have formed around the outside, as in Fig. 24-11. As a result, lava flows are often left with a hollow tube through them. Occasionally the tops of these tubes cave in and we can explore the length of the tubes, some of which extend for more than a mile through a flow. We may find ice in a few lava caves from New Mexico northward, even when the weather outside is quite warm. See if you can work out an explanation for the occurrence of this ice from what you already know of the behavior of air and water.

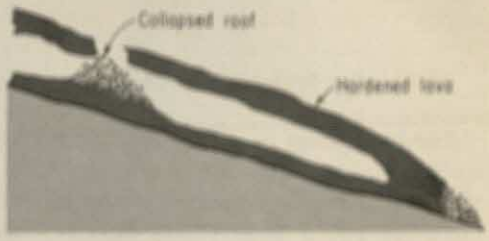


Fig. 24-11. Cross-section of a lava tube or cave.

24-10 Volcanic Activity in Continental United States

The only large-scale volcanic activity that has taken place within recorded history inside the continental United States was the eruption of Mt. Lassen (Fig. 24-12) in northern California in 1914–15. During this eruption, a very large blast blew out horizontally from beneath the lava plug of the volcano, causing great destruction of plant and animal life in the immediate vicinity but no loss of human life. Mt. Lassen has several thermal areas associated with it in which hot springs and fumaroles abound, but there seem to be no signs of recurring activity of the mountain itself. During the days of the Gold Rush, many reports circulated of a large fire that lasted for a number of days to the north in the vicinity of Lassen Peak. Since there are recent lava flows and a cinder cone in the region described, we may speculate that it was the eruption of one of these cinder cones that was reported by the early California miners.

In the state of Washington, fairly substantial records report activity on the part of Mt. Baker in the 1850's. Mt. Rainier, which has a crater nearly half a mile in diameter, issues jets of steam from vents in its crater, and it too is reported to have been active in pioneer days around 1842. Sailors coming into the mouth of the Columbia River about 1841–42 accurately reported an eruption of Mt. St. Helens that occurred at that time. Mt. Hood also has a record of activity in 1859 and 1865, and even today mountain climbers and skiers often notice the odor of sulfur fumes near Crater Rock on the mountain. There are numerous lava flows in New Mexico, Arizona, California, Utah, Oregon, and Idaho. A number of our national monuments and national parks were established to preserve the results of volcanic activity,

Fig. 24-12. Mt. Lassen, California, the only active volcano in the continental United States, shown here erupting in 1915. (World Wide Photos)

which stand as reminders of how short a time has elapsed since our western states were an actively volcanic region.

24-11 Hot Springs and Geysers

The hot springs in this country are located predominately in the western states. There are over 1000 such springs in the West, 46 in the Appalachian Highlands of the East, 6 in Arkansas, and 3 in the Black Hills of South Dakota. The western thermal springs commonly get their heat from contact or close association with hot magma. As ground water descends, the thermal gradient is quite high owing to the presence close to the surface of remnants of hot intrusive rocks. The heated water rises and appears as a spring, geyser, or fumarole. Some of the springs in the eastern United States result from a special condition of ground water circulation. Apparently the ground water can find a path deep enough into the earth to be heated by just the normal rise in temperature that is found in the earth's crust. At Warm Springs, Georgia, the water is thought to seep downward to a depth of 3800 feet, and it is warmed approximately 1°F every 100 feet it descends. The warmed water becomes less dense than the descending water and therefore rises to the surface where it emerges as hot springs.

When the supply of water is insufficient to carry away the heat that is acquired, the water may turn to steam and produce a steam vent or a fumarole. If the amount of water is sufficient to carry away the heat but not enough to overflow the outlet basin, the hot water may dissolve the sides of its basin and result in a cloudy or turbulent spring. As the water decomposes more of the rock material around the sides of this spring, it becomes more fully charged with rock dust and in an extreme case may produce a mud pot. If by any chance the ground water becomes charged with carbon dioxide through contact with volcanic gases, it may appear as a carbonated spring. When this hot carbonated water rises through limestone layers, it will readily dissolve the limestone and bring the dissolved calcium carbonate to the surface where the water will evaporate, lose its load of dissolved material, and build up masses of rock known as *travertine*. This is the origin of Mammoth Hot Springs in Yellowstone National Park.

Of greatest interest among hot springs, however, are the periodically erupting ones that are known as *geysers*. If the hot carbonated water rises through rock containing large quantities of silica, the deposited material that builds up near the mouth of the spring is hard rock called *geyserite*. Because this rock is harder than travertine it can withstand the periodic explosions of hot water and steam. Geysers occur on a large scale in only three regions of the world—Yellowstone National Park, Iceland, and New Zealand. There are great varieties of geysers and geyser activities; some

erupt only a few feet, whereas one in New Zealand once threw water to a height of approximately 1000 feet. Probably the most famous geyser is Old Faithful in Yellowstone, which erupts to a height of 110 to 160 feet and blows out about 10,000 gallons of water during each eruption. Its regularity and beauty have endeared it to millions of tourists.

The mechanism of a geyser eruption is not simple, but we can understand much of it from what we have already learned about the heat properties of water. We know that as the pressure on water is increased with depth, the boiling point of the water becomes much higher than it is on the surface. When the geyser tube is tortuous, so that the water cannot circulate freely in the tube, the temperature rises much faster near the bottom of the water column where heat is applied than at the top (Fig. 24-13). Eventually the water approaches the boiling point even at the high pressures existing at the depths involved. As the water is heated it expands and some may spill out of the top of the geyser column, causing the pressure on the remaining water to be lowered. With the lowering of pressure, the water finds itself suddenly above the boiling point and much of it turns to steam, the sudden expansion of which pushes large volumes of very hot water out of the vent. Some of this water, too, may turn to steam as it finds itself at a much reduced pressure. The resulting eruption throws water and steam to varying heights above the land surface. After the water has expanded as steam and is cooled by contact with the air, it falls back into the geyser tube, is further cooled, and, trickling down the tube, causes the steam there to condense and leave a partial vacuum in the geyser chambers. Water from the surface and ground water then enter the geyser chambers under the force of atmospheric pressure, and the heating process starts all over again.

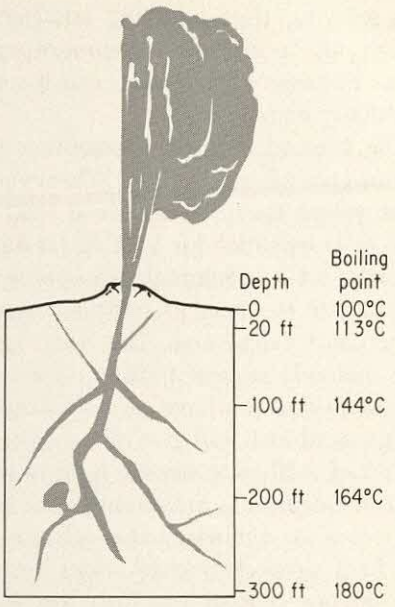


Fig. 24-13. Diagram of a geyser.

24-12 Heat in Biological Processes

Although we have been primarily concerned with heat in our physical environment, we have also found a number of places in the biological world where heat plays a vital role. A brief summary may help us to put them in the right perspective. In Chapter 20 we noted that the body must have a well-regulated and uniform temperature in order for the rate of chemical activity to be constant. The oxidation of food within the cells of our bodies is the source of this heat, and since the skeletal muscles constitute about one-half of the active tissue within the body and since they are the most active of the tissues, it follows that most of the heat is generated within these muscles. In our warm-blooded bodies, when the temperature drops to a colder level, the rate at which we use food (metabolism) is speeded up, and more heat is generated. If our body is not able to provide sufficient heat in other ways, we may shiver from increased

muscular activity, thus releasing additional heat. This act of shivering may increase the amount of oxidation within the body by as much as 400 per cent, and it serves as a powerful heat-producing activity.

We have learned, too, that sometimes it is necessary for the body to give off excess heat. Whenever we are in an environment where the temperature is near or above that of our body, it is essential for heat to be eliminated from our body in order for it to maintain a uniform temperature. Failure to eliminate the heat in sufficient quantities may result in fever or heat exhaustion. The body has to make use of the three methods of heat transfer—conduction, convection, and radiation—that we have already studied in Chapter 20. An average-sized man will give off an amount of heat equivalent to that of a 60-watt electric light bulb while he is resting. Most of this heat is lost through the skin (85 per cent), but some of it is eliminated through the lungs and by excretions. In a properly heated room during the winter, as much as 50 per cent of the heat loss is accomplished by radiation, 25 per cent is lost by convection, and another 25 per cent by evaporation.

Most of us know that when our bodies are exposed to a cold environment, the blood does not flow as freely through the skin, and we all have seen the red face or flushed appearance of a person when the temperature is high. As our body temperature rises, the blood vessels in the skin dilate while those in the abdomen constrict, and the blood is forced to flow in larger quantities to the skin, where the blood is cooled by radiation and convection. On a cold day, the reverse takes place, and the warm blood is restricted in its circulation through the skin where it would lose heat readily.

We have already investigated the importance of the evaporation of perspiration from the skin's surface as a means of heat transfer. Convection and conduction also play an important role in the control of body heat. As the relative humidity is increased, the amount of heat transferred to or from the body by these methods is also increased. At or below 65°F there is very little perspiration, and the body is cooled even when the humidity is high. For this reason, we find it easier to withstand cold weather when the amount of moisture in the air is small. Perspiration usually starts at temperatures above 70°F, and an excessive humidity decreases the amount of evaporation and consequent heat loss from the skin. Between these two temperatures—65°F and 70°F—the relative humidity appears to have little effect upon the heat loss by the body. The production of heat and the regulation of the body temperature is a delicate and complicated process involving many organs and tissues of the body. Although it is not well known, heat is also produced in the bodies of cold-blooded animals and in plant tissues by the process of oxidation. In plants the heat is usually dissipated quickly to the surroundings.

We can summarize the role of heat in biological processes by stating that all oxidation of food results in the release of heat energy. This heat tends to raise the temperature of the living organism and necessitates, in warm-blooded animals, a careful and complex mechanism for bodily temperature control. This heat energy that is released provides the thermal environment in which the proper degree of chemical activity essential for life can take place, and we might say that heat energy is the energy of life.

24-13 Summary

The temperature at the earth's center, according to our most educated guess, is about 5000°C. Geologists generally believe that the earth's heat is derived from several sources—the original heat left from the formation of the earth, the radioactivity of the earth's materials, and the compression of the earth's interior due to the gradual cooling and shrinking of the earth's crust.

Molten rock found within the earth is called magma, but if it appears at the surface it is known as lava. When the magma forces its way into and through cracks in other rock structures it is called intrusive activity, and the resulting rocks are known as intrusive formations. If the magma forms a horizontal layer, this layer is called a sill, but if the formation forms a vertical structure cutting through the rock, it is known as a dike. Large bodies of intrusive rock may form a batholith if they have no apparent lower floor and extend over a large enough horizontal area; a smaller intrusion, which has a lower floor and succeeds in pushing up the overhead rock structures into large domes or even mountains, is known as a laccolith.

As molten rock reaches the earth's surface, the decrease in pressure allows any dissolved gases to expand and escape, resulting in the formation of an explosive volcano or a pyroclastic or cinder cone. If the magma has a small amount of dissolved gas, the flow from a vent may slowly build up a shield volcano around the central vent. A combination of quiet flow and periodic explosions produces a composite cone. When lava is fluid enough to flow quickly from a fissure of some length, a fissure flow results. If a volcano suffers a major eruption, a large circular pit known as a caldera may remain.

A volcanic eruption may eject so much dust into the atmosphere that it profoundly affects the world's weather for months or years. Small volcanic debris is known as lapilli; larger pieces may take a streamline shape and are known as volcanic bombs. If the volcanic rock contains large amounts of dissolved gas and solidifies quickly, trapping the gas in tiny bubbles within the rock, a very light, porous material results, which is known as pumice. If a volcanic erup-

tion causes large amounts of debris to be blown downward so that the rocks are "lubricated" by the escaping gas, an unsorted mixture of rocks known as *nuée ardente* results. Volcanic activity has shaped our western states by raising large volcanic peaks and many smaller cinder cones, by covering many hundreds of square miles with lava flows, and by making possible hot springs and geysers.

But heat in natural processes is not limited to the weather, climate, volcanic activity, and all the physical aspects of our environment. It includes as well the very complex processes of oxidation and heat control of living beings.

EXERCISES

A. KEY TERMS TO REMEMBER

batholith
caldera
composite cone
dike
extrusive
fissure flow
fumarole
geyserite
intrusive
laccolith
lapilli
magma
metamorphic rocks
nuée ardente
pumice
pyroclastic cone
shield volcano
sill
travertine

B. QUESTIONS ABOUT HEAT IN NATURAL PROCESSES: II

1. In Fig. 24-2, which has the higher specific heat—the solid state just before melting or the liquid state just after melting? Is the specific heat of the solid state ever the same as that of the liquid state in Fig. 24-2?

2. Throughout the temperature range indicated for the solid state in Fig. 24-2, is the specific heat uniform?

3. Plot on a map of your state the location of evidences of the earth's heat.

4. On an outline map of the United States, locate and name each of the national parks and national monuments which are in some major way related to the earth's heat.

5. In Fig. 24-7, which of the rocks indicated were formed in the deepest water?

6. Determine the probable order of the geological events represented in Fig. 24-4.

7. The following is an improper definition of a volcano: "A volcano is a burning mountain from the top of which issues smoke and fire." State in each of the following whether it is true or false:

(*a*) No "burning in the sense of combustion, such as the burning of wood, occurs in a volcano; (*b*) volcanoes are not necessarily mountains; (*c*) activity takes place more commonly on the sides or flanks than at the summit; (*d*) the "smoke" is condensed steam often mixed with dust particles until dark colored; (*e*) the "fire" is the reflection of the red hot material on the vapor clouds above the volcano.

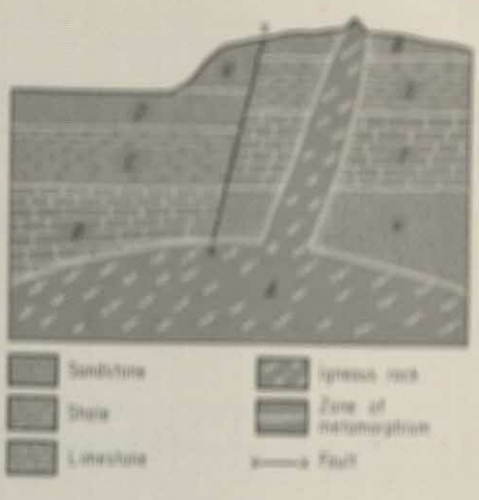


Fig. 24-14. A geological structure.

Key A

A. Event *X* occurred before event *Y*.
B. Event *Y* occurred before event *X*.
C. Events *X* and *Y* occurred at approximately the same time.
D. Event *X* and/or event *Y* did not occur.
E. There is insufficient information to warrant any of the above answers.

Key B

A. In a sequence of sedimentary rock strata, the lowermost layer is the older.
B. Diastrophic movements resulting in folding have occurred later than the youngest bed folded.
C. Diastrophic movements resulting in faulting have occurred later than the youngest bed faulted.
D. An intrusive igneous rock is younger than the youngest bed intruded.
E. None of the above.

Select the answer to the following items from Key A:

Event *X*	Event *Y*
8. formation of C	formation of B
9. formation of A	formation of B
10. formation of XX	formation of A
11. formation of B	formation of Y

Select the answer to the following items from Key B:

12. Which principle in key B serves as the basis for deducing the answer to question 8?

13. Which principle in key B serves as the basis for deducing the answer to question 9?

14. Which principle in key B serves as the basis for deducing the answer to question 10?

15. Which principle in key B serves as the basis for deducing the answer to question 11?

16. In Fig. 24-14, was the land rising or falling while T and S were forming?

17. Do you know of any effect that volcanic activity has had on music? literature? art? the beliefs of men?

18. Why is the basement of a house usually cool during the summer?

19. What properties of water make it especially suitable as the major fluid in a living organism?

20. What properties of water make it especially suitable as a major fluid in an environment in which life is to exist?

21. From Table 24-1 and your experience in chemistry, what do you consider the most significant difference in the composition of the lavas to give them their "acid" or "basic" structure?

C. PROBLEMS ABOUT HEAT IN NATURAL PROCESSES: II

1. Verify that extrapolation would lead to a temperature of 350,000°F at the earth's center.

2. What must be the average temperature gradient in the earth if the temperature at the center is 5000°C?

3. Old Faithful throws out 10,000 gallons of water at each eruption. If this water were raised from 50°F to 180°F in 60 minutes, how many Btu's are transferred each minute? (1 gal = approximately 8 lb of water.) Ans. 173,000.

4. The lava flow that began two days after the start of the Paracutin eruption flowed at what average rate (ft/sec) for the first 50 days?

Twenty-Five

Radioactivity

In Chapter 12 we saw briefly how the discovery of radioactivity provided the key to open the door to the inside of the atom. At that time we were chiefly concerned with the structure of the outer parts of the atom—the energy levels of the electrons and their effect upon the valence and other chemical properties of the elements. We investigated briefly the general constitution of the nucleus of the atom and saw that all the protons and neutrons are bound together in the tiny core of the atom. However, in order to appreciate the nature of the tremendous sources of energy locked within the atom, we must learn more about the nucleus. Let us review our earlier discussions of radioactivity and atomic structure and extend these remarks to a more detailed examination of the atom's nucleus.

The modern picture of the atom reveals a strange domain. In the extremely small world of atomic dimensions, we have already encountered negatively charged particles (electrons), positive charges (protons), and particles having no charge (neutrons). The discovery of these particles spans the years from 1897 to 1932, from the early work of Crookes, who discovered cathode rays, through the work of scores of scientists to Chadwick's discovery of the neutron. But much more is now known about the particles that compose the atom. In addition to the ones named above, several other particles have been found that are important to an understanding of the nuclear atom.

The electrons in the outer parts of the atom occupy definite energy levels. Each energy level may be occupied by a certain number of electrons—two in the lowest level, eight in the second, eighteen in the third level, and so on. We found

that electrons may be elevated to higher energy levels if they are excited with sufficient energy, after which these electrons fall back to their stable levels. The energy that is released when these electrons return to their original levels appears as light quanta or photons having definite frequencies. It was through a study of the light emitted by these electrons, known as atomic spectra, that the various energy levels of the electrons were determined. This theory was called the quantum structure of the atom (p. 232).

25-1 The Nuclear Atom

The nucleus of the atom, we learned, contains all the protons and neutrons of the atom. The arrangement and behavior of these nuclear particles provide us with one of the newest frontiers in physical science. The fact that an atom of, say, gold contains 79 protons and 118 neutrons in a nucleus that is only 10^{-12} cm in diameter may cause us to wonder. How can so many positively charged particles exist in so small a space? If they are typical positive charges, they should repel each other mightily at such close ranges. Yet we know that the gold atom is very stable.

There is general agreement among scientists that the particles in the nucleus of the atom occupy energy levels and that these nuclear energy levels are analogous to the electronic energy levels in the outer part of the atom. When there is a change between energy levels in the nucleus, the energy that is emitted is much greater than when electrons change levels. An electron when shifting energy levels may change in energy only a few electron volts, but energy changes in the nucleus are typically as large as a million electron volts (mev). From Equation 12.1, $E = hn$, we learned that the energy E of a photon is proportional to the frequency of the wave associated with the photon. A typical electron in making an energy level transition gives off energy in the visible region of the electromagnetic spectrum. The energy changes in the nucleus, however, are possibly a million times greater. Thus, the photons emitted in nuclear changes have frequencies in the range of a million times that of visible light. Such frequencies may have wave lengths as short as 1.2×10^{-12} cm, corresponding to very short X rays, and these rays—called gamma rays—are, of course, the extremely penetrating ones that are emitted in radioactivity. Gamma rays, therefore, serve to give us a measure of the energy changes that occur in the nucleus of the atom.

The nucleus contains other particles besides protons and neutrons. In 1934, H. Yukawa of Japan proposed that the strong repulsive forces in the nucleus must be balanced by similar attractive forces, and he calculated that the nucleus must contain a negatively charged particle with a mass about 150 times that of the electron. Two years later, such a particle was actually discovered. It was called a *meson*,

at the time, but it is no longer classified as such. Since then several kinds of mesons—negative, positive, and neutral ones—have been found having masses slightly more than 200 times that of the electron. The *positron*, discovered by Anderson at the California Institute of Technology, is a positively charged particle that is essentially the counterpart of an electron. Figure 25-1 shows the paths of an electron and a positron in a cloud chamber, an instrument described on p. 514. That the two paths show similar but opposite directions of curvature indicates that they have the same mass but opposite charges. The properties of some of these, along with those of other fundamental particles, are listed in Table 25-1.

Many other mesons and other particles have been discovered, but their consideration, although interesting, lies beyond the scope of this book.

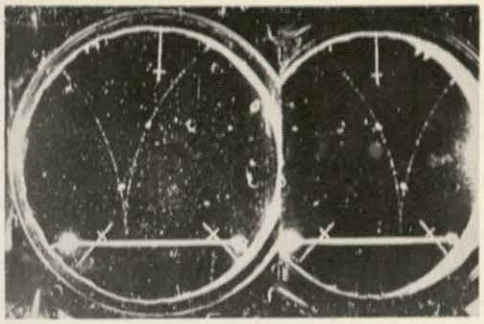


Fig. 25-1. Cloud chamber paths produced by an electron and a positron. They originate near the bottom where a high-energy gamma ray has struck a thin sheet of lead. (Courtesy W. A. Fowler, California Institute of Technology)

TABLE 25-1 SEVERAL IMPORTANT SIMPLE PARTICLES

(Masses are compared to the mass of the electron)

Positive particles	*Neutral particles*	*Negative particles*
Proton Mass = 1836	Neutron Mass = 1838	Antiproton Mass = 1836 (?)
pi+ meson Mass = 273	pi^0 meson Mass = 264	pi− meson Mass = 273
Positron Mass = 1	Neutrino Mass = 0	Electron Mass = 1

25-2 Natural Radioactivity

Following the discovery in 1898 by Marie Curie of the radioactive element, polonium, about fifteen natural elements have been found to be radioactive. The products of disintegration of all these elements are identical—the alpha particle, the beta particle (an electron), and gamma rays. There are, however, considerable differences among these elements in the manner or intensity of their radiations. Some elements give off alpha and beta particles quickly, whereas others decay only slowly. Furthermore, the energy of the emitted particle varies over a range and gives what might be called an energy spectrum. For instance, radium emits an alpha particle at 4.79 mev, while thorium emits an alpha particle at 7.20 mev. As a result of these two factors—the speed of decay and the energy of the emitted particle—the relative intensity of radioactive emission may vary billions of times between the most active and the least active radioactive element.

Radioactivity is the spontaneous decay of the atomic nucleus. There are no known means of speeding it up or slow-

ing it down. If we were to examine a sample of a radioactive substance and measure the quantity or intensity of radioactivity that it shows over a period of time, we would find that it gradually becomes less radioactive. A careful look at the curve in Fig. 25-2 reveals an important concept in radioactivity. In all radioactive processes the intensity of radioactivity falls off by what are known as half-life periods. A *half-life period* is the length of time for the intensity of radioactivity to fall to half the original amount. Thus, if we start with one gram of the radioactive element, after the half-life period only half a gram will remain. After the second half-life period, a quarter of a gram will remain, and at the end of the third half-life period only one-eighth of a gram will remain. The half-life period of radium is 1590 years, and the half-life period of uranium is 4.5 billion years. This difference in half-life periods explains the difference in the intensity of the radiations emitted by these two elements. The half-life periods of other radioactive elements are noted in Fig. 25-3.

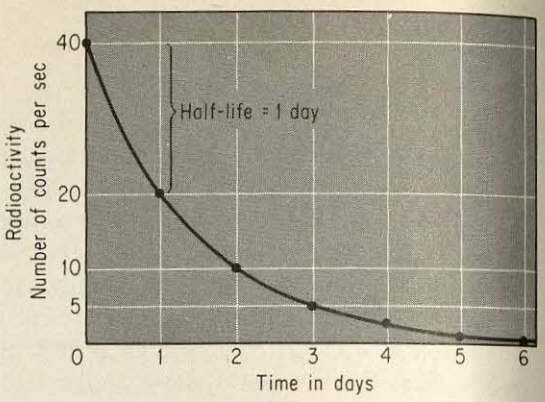


Fig. 25-2. The decay of a radioactive substance.

An element that emits an alpha particle in effect loses two protons and two neutrons from its nucleus. The resulting atom has an atomic number two less than the original atom, and its mass is four units less. An element that loses a beta particle from its nucleus gains one in atomic number. This gain may be accounted for in this way: since there are no electrons in the nucleus, the beta particle is derived from the neutron, thus forming a proton. What actually occurs is one of the great mysteries in physics, but for our present purposes we might say that the process involves the hypothetical reaction:

$$\text{Neutron} \rightarrow \text{proton} + \text{beta particle} + \text{neutrino}$$

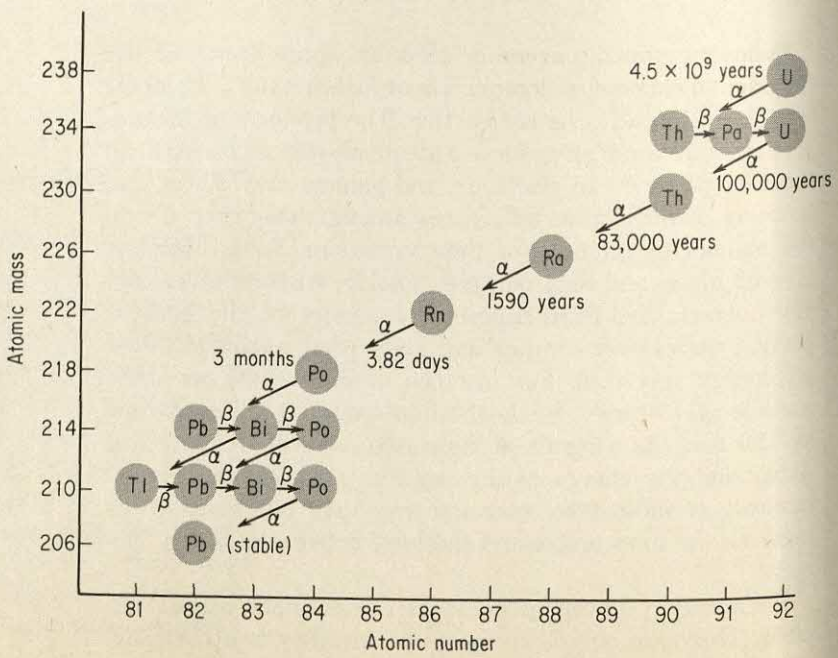


Fig. 25-3. The uranium-238 radioactive decay series.

Applying these concepts to uranium-238, let us follow the process of radiactive decay that eventually produces the stable element, lead. Figure 25-3 gives a detailed account of this process. U-238 first loses an alpha particle, then two betas, then successively five alphas, two betas, an alpha, two beta particles, and finally an alpha particle. The end product of this series of decay reactions is lead-206, a stable element.

In addition to the uranium-238 series of radioactive decay, there are three other similar series, each of which ends with a stable element, usually lead. Apparently elements having atomic numbers greater than 82 and atomic masses greater than 206 have unstable nuclei. This means that a nucleus that contains more than 82 protons and 124 neutrons undergoes spontaneous disintegration, losing alpha and beta particles until a stable nucleus is reached.

It is interesting to note that there is a rather narrow range of atomic numbers and atomic masses that give elements with stable nuclei. This unusual characteristic is noticeable when we examine Fig. 25-4, which shows the isotopes of the elements that occur in nature. Very few of the lighter elements are radioactive, yet all of the heaviest elements are radioactive, and the unstable isotopes among the lighter elements have a greater number of neutrons than do their stable cousins. There seems to be a certain regularity in the ratio of pr tons to neutrons in stable isotopes. You will recall that in Chapter 13 we found that isotopes having even atomic numbers and even numbers of protons are the most plentiful. Similarly, we find here that the most abundant isotopes are possessed by those elements of even atomic numbers, while the odd-odd isotopes are extremely rare.

25-3 Nuclear Stability

Now let us turn our attention to the problem of nuclear stability. What holds the nucleus together, why are certain

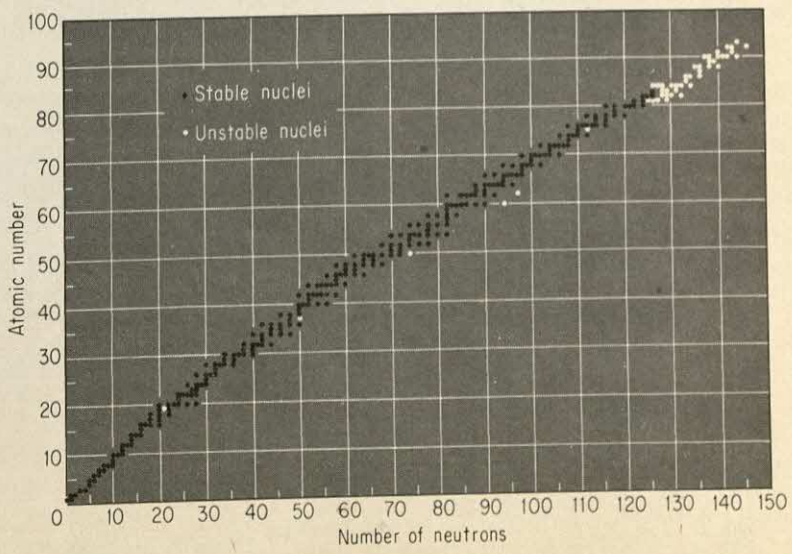


Fig. 25-4. Stable and unstable isotopes of the naturally occurring elements.

isotopes unstable, and how much energy is involved in the decay of these unstable nuclei? First, what about the energy involved in the simple atoms of hydrogen and helium?

A hydrogen atom, having a mass of 1.0081, contains a single proton in its nucleus. A neutron has a mass of 1.0090. If we hypothetically combine two hydrogen atoms with two neutrons we have the makings of a helium atom, the atomic mass of which is 4.0039. But if we add up the masses of these four particles that we have combined, we find a discrepancy:

2 hydrogen atoms	$2 \times 1.0081 = 2.0162$
2 neutrons	$2 \times 1.0090 = 2.0180$
Total mass	4.0342
Mass of helium atom	4.0039
Difference	0.0303

Here we find that there is a loss of 0.0303 mass unit per atom of helium formed from simple particles. Is there any way we can relate this mass loss to the energy of the nucleus of the helium atom?

In 1905, Einstein developed his famous equation which showed the relationship between mass and energy. This equation, derived from purely theoretical deductions, has been experimentally proved in recent years, and we can therefore be reasonably assured of its validity. The Einstein equation is

$$E = MC^2 \qquad (25.1)$$

energy = mass times the square of the velocity of light

In this equation, when the mass is expressed in kilograms and the velocity of light is expressed in m/sec (3×10^8 m/sec), the energy is in units of joules. Thus, for each gram-atomic weight of helium formed from hydrogen atoms and neutrons, 0.0303 gram of mass is converted into energy:

$$E = 0.0000303 \text{ kg} \times (3 \times 10^8 \text{ m/sec})^2$$

$$E = 2.73 \times 10^{12} \text{ joules} = 6.52 \times 10^{11} \text{ cal}$$

This is an enormous amount of energy—millions of times that released in the most vigorous chemical reactions.

From the above deduction, we may expect the helium atom to be much more stable than the two hydrogen atoms and two neutrons of which it is made. In fact, it is more stable by 2.73×10^{12} joules. This quantity of energy expressed in electron volts is equivalent to 28 million electron volts (mev). The energy, then, that holds the two protons and two neutrons together in the helium nucleus is 28 mev, or 7 mev per nuclear particle. This 7 mev is called the *binding energy* of each particle in the nucleus of the helium atom—and is a measure of the energy that must be expended in breaking the attractive bonds in the nucleus.

If we consider the nucleus of the oxygen atom, composed of 8 protons and 8 neutrons, we find a still higher binding energy. Hypothetically, the nucleus of the oxygen atom may

be made of 8 hydrogen atoms and 8 neutrons, the total mass of which adds up to 16.1360. Since the actual mass of the oxygen atom is 16.0000, the mass defect is 0.1360 mass units. This mass, when converted to its equivalent energy and divided among the 16 particles in the nucleus, gives a binding energy of 8.0 mev per nuclear particle. In a similar way, the binding energies of all the elements can be found. Figure 25-5 shows the binding energy for many of the elements, and from this curve we may conclude that the elements from approximately mass 40 to 120 have the highest binding energy and should have the most stable nuclei. Since this range includes calcium, iron, nickel, and many other stable elements, our conclusion is dependable. On the other hand, we may expect to find those elements at the extreme ends of the curve to have the most unstable nuclei. Consequently, the lightest elements, hydrogen, helium, and lithium, can be expected to give up large amounts of energy when they fuse together to form heavier nuclei, and the heaviest elements release a great deal of energy when they undergo radioactive decay or other types of nuclear changes.

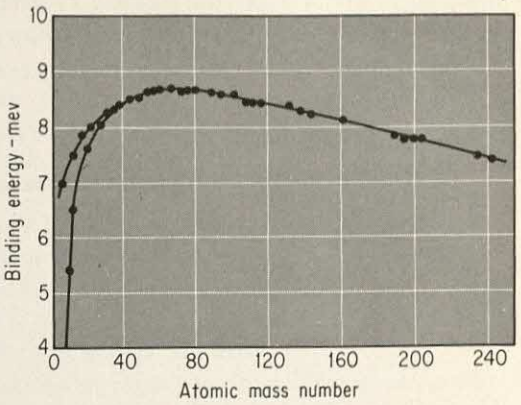


Fig. 25-5. The binding energy of the nuclear particles of the various elements.

Another way in which we may view the stability of the atomic nucleus of an atom is in terms of potential energy. The nucleus of the atom contains protons and neutrons, which, through some mysterious force, are tied together very securely in the nucleus of a stable atom. We might visualize these particles as existing in an energy well. The high walls of the energy well are a barrier to the escape of any particle from the nucleus (Fig. 25-6a). In a radioactive element, the particles in the nucleus exist at higher levels in the energy well of the nucleus. Consequently, a nuclear particle occasionally passes the energy barrier at the boundary of the nucleus, and after a series of such losses of high-energy particles from the nucleus, the energy of the remaining particles subsides to a stable level in the energy well.

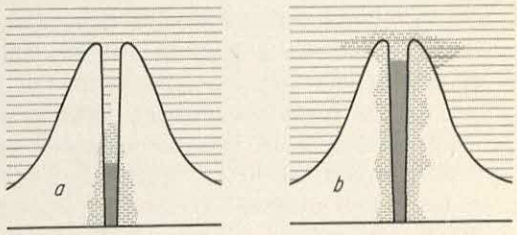


Fig. 25-6. Potential energy barriers of (a) stable nuclei and (b) unstable nuclei.

Still another characteristic of the stable atomic nucleus is shown by those isotopes that are most abundant in nature. We find that among the most stable isotopes occurring in nature (including those in the sun) are those of helium, oxygen, calcium, nickel, tin, barium, and lead. Of these and similar elements, the most stable ones are those that contain either 2, 8, 20, 28, 50, 82, or 126 neutrons or protons. There appears to be something mysterious about these numbers. They point toward some kind of relationship within the nucleus about which we may merely make some reasonable guesses. One of these guesses is that these numbers represent completely filled energy levels, like the electronic energy levels in the outer part of the atom.

25-4 Artificial Transmutation

Ever since ancient times, man has sought the means of converting one element into another. History records many un-

successful attempts to convert lead to gold, but these early efforts were doomed to failure because they were based upon chemical reactions in which the energy changes are far too small to produce nuclear changes. We know now that, in order to change one element into another, we must change the number of protons in the nucleus of the atoms of the element. This requires considerable energy because of the large energy barrier at the boundary of the nucleus.

Suppose we wish to convert nitrogen to oxygen, which requires the addition of a proton to the nucleus of the nitrogen atom. Since high-energy alpha particles are easily obtained from some of the naturally radioactive elements, we can aim alpha particles at the nitrogen atoms in the hope of having a reaction. We will soon find that this is not a very efficient process. As the positively charged alpha particles approach the nucleus, they are severely repelled by the positively charged nucleus. Only in one of several million tries will we succeed in getting an alpha particle to strike close enough to surmount the energy barrier (Fig. 25-7). The shape of this energy barrier obviously presents an extremely small target, but when at last an alpha particle falls into the nuclear energy well, a reaction occurs as follows:

$$_{7}N^{14} + {}_{2}He^{4} \rightarrow {}_{1}H^{1} + {}_{8}O^{17}$$

This equation represents the transmutation that was the first successful artificial change of one element into another. It was accomplished by Lord Rutherford of England in 1919. The equation says that nitrogen, with an atomic number of 7 and a mass of 14, when bombarded by alpha particles yields a proton (at. no. = 1; at. mass = 1) and an oxygen isotope with a mass of 17.

Many transmutations have now been accomplished. In addition to alpha and beta particles from radioactive elements, we now have other particles that can be elevated to extremely high energies (high velocities). One of these is the *deuteron*, the charged nucleus of an isotope of hydrogen that contains a proton and neutron. With these high-energy particles, even the heaviest elements can be successfully transmuted to higher elements. Some of the instruments used to accelerate these particles will be described in the next section. Nuclear changes may be brought about by these particles:

alpha particles	$_{2}He^{4}$
protons	$_{1}H^{1}$
deuterons	$_{1}H^{2}$
neutrons	$_{0}n^{1}$

Following are some of the more important nuclear transmutations that have been performed. When beryllium is bombarded with alpha particles, it is changed to carbon:

$$_{4}Be^{9} + {}_{2}He^{4} \rightarrow {}_{6}C^{12} + {}_{0}n^{1}$$

This reaction is the one by which Chadwick identified the neutron in 1932.

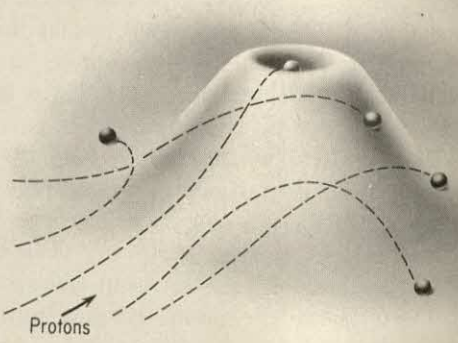


Fig. 25-7. The potential energy barrier of the nucleus toward a positively charged particle.

Silicon may be produced from aluminum:

$$_{13}Al^{27} + {}_{1}H^{1} \rightarrow {}_{14}Si^{28} + \text{gamma ray}$$

and a radioactive isotope of sodium is commonly made by bombarding magnesium with deuterons:

$$_{12}Mg^{24} + {}_{1}H^{2} \rightarrow {}_{11}Na^{22} + {}_{2}He^{4}$$

Again, an isotope of an element may be converted into another isotope of the same element by neutron bombardment. There have been hundreds of radioactive isotopes produced artificially, ranging in atomic mass from the light to the heavy elements. Many of these artificially radioactive isotopes have proved to be very useful in studying problems that have stumped scientists for a long time. We shall refer to some of these applications later.

25-5 The Acceleration of Charged Particles

In order to bring about the transmutation of many isotopes, particularly the heavier ones, particles that possess extremely high energy are required. It is necessary to have protons, deuterons, alpha particles, and electrons accelerated to velocities sufficient to penetrate the nuclear energy barrier, and this is accomplished by electromagnetic devices, the names of which in recent years have become common. One such instrument is the cyclotron, invented by E. O. Lawrence in 1929 at the University of California.

The *cyclotron* operates on two simple principles. One of these is that a particular charged particle or ion is attracted to an electrical pole of the opposite sign. The other principle is that a charged particle is accelerated in a magnetic field in a direction perpendicular to the direction of the magnetic field. Both of these principles were discussed in Chapter 9. Although the apparatus for controlling these elec-

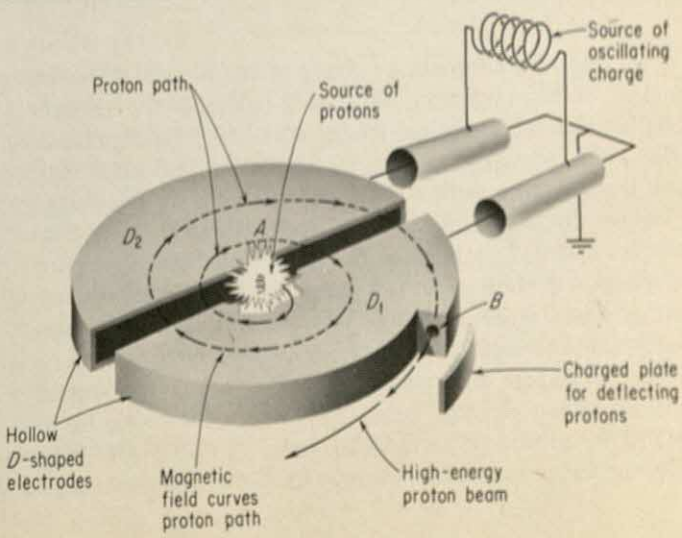


Fig. 25-8. The principle of the cyclotron.

trical and magnetic fields is extremely complex, the principle of the operation of the cyclotron is simple.

In Fig. 25-8 a cutaway view of the magnetic pole pieces of the cyclotron is shown. Charged particles, say protons or deuterons, are admitted at *A* into the heart of the device, which consists of two hollow D-shaped electrodes. The charged particles are first attracted toward the positive *D*, then the polarity of the dees (D_1 and D_2) are changed alternately. The charged particles move back and forth in the evacuated space inside the dees. At the same time, a strong magnetic field causes these particles to move in a horizontal circular motion, that is, perpendicular to the magnetic field. Moving faster and faster, these particles gain more and more velocity as they move around in spiral paths. The time for the ions to make a complete turn around the dees is the same, regardless of the diameter of its circular path. The dees are alternately charged positively and negatively perhaps ten million times a second. Hence a stream of ions can be taken in at *A* and made to undergo continuous acceleration. Eventually, at a small port at *B*, a positively charged plate pulls the stream of high-energy ions from its circular path. These high-energy particles may then be used for bombarding nuclei (Fig. 25-9).

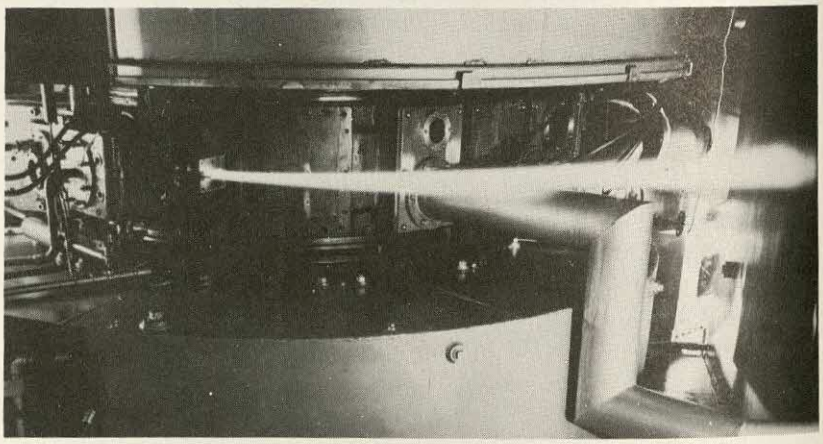


Fig. 25-9. A beam of high energy particles emerging from a 60-inch cyclotron. (Courtesy University of California Radiation Laboratory)

If we apply an alternating voltage of 20,000 volts between the dees of the cyclotron, with each half-turn the particle (ion) is given an additional energy equal to its charge times 20,000 volts. After the ion has made 250 revolutions it may leave the chamber with a total energy equal to its charge multiplied by 20,000 $\times$ 250 $\times$ 2 or 10,000,000 volts. The quest for charged particles with higher and higher energies has led to the construction of larger cyclotrons, as well as to other kinds of accelerating instruments. One of the most effective and elaborate of these is called the *bevatron,* a machine designed to produce charged particles possessing up to 31 billion electron volts (bev, hence the name bevatron). The instrument at the University of California (Fig. 25-10) provides a circular pathway for the charged particles

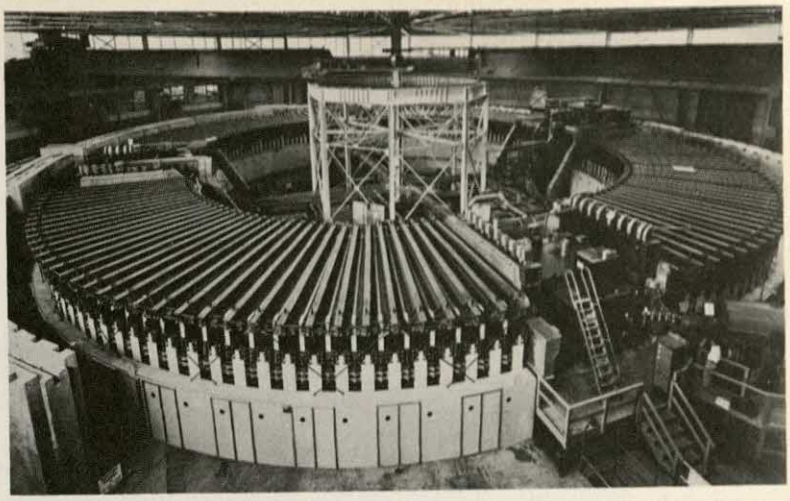


Fig. 25-10. The bevatron. The size of this instrument is indicated by the man standing in the lower right of the picture where the charged particles are injected into the bevatron. (Courtesy University of California Radiation Laboratory)

that is 400 feet in circumference. Each time the charged particles, such as protons, make a circuit, they are accelerated about 1300 electron volts. After 4,000,000 times around the circle, or a distance of 300,000 miles, the protons have enough energy to strike target atoms and produce mesons. The bevatron provides the means of studying more intimately these mesons, which are believed to be the key to an understanding of the tremendous forces that bind atomic nuclei together. Here, indeed, is an "atom-smasher."

In addition to the cyclotron and bevatron, there are several other instruments that have been designed for the specific purpose of accelerating charged particles. These include the Van de Graaf generator (primarily an electrostatic generator) used in conjunction with a linear accelerator, the betatron (which produces both high-energy electrons and extremely penetrating X rays), and the synchrotron (which is a somewhat modified cyclotron). All of these instruments operate on basic principles of electrostatics, electricity, and magnetism, although we should not overlook the high degree of ingenuity required to design and operate them.

25-6 The Measurement of Radioactivity

Keeping pace with the rapid development in artificial transmutations are the methods of detecting and measuring radioactivity. The earliest method of detecting alpha particles—Becquerel's photographic plates—has become even more popular in recent years. All charged particles, including protons, electrons, and mesons, affect the photographic emulsion. The sensitivity of common photographic film is increased by providing a much greater amount of silver bromide in the emulsion than ordinarily. Advantages of the photographic plate detection process are its simplicity, its continuous sensitivity, and the permanent record it makes of the characteristic paths of these particles.

Other methods of detecting alpha, beta, and gamma rays rely primarily upon the ability of these high-energy particles to produce ionized molecules in the gas through which they pass. For example, an alpha particle emitted from a radio-

active source may produce 100,000 ions per centimeter of ordinary air through which it travels. In effect, this renders the air a conductor at the region where the alpha particle traveled. If an electrical circuit could be established that depends on the conductivity of the air as affected by an alpha particle, the presence of the alpha particle could be detected. Such is the principle of the Geiger-Müller tube.

The *Geiger-Müller counter* is essentially a means of measuring the number of ionized paths formed by radioactive rays. When a beta particle passes through the window of the counting tube (Fig. 25-11), it leaves a trail of ionized gas molecules. The resulting electrons that are stripped from the gas molecules by the alpha particle form an avalanche of electrons that moves toward the positive electrode in the counting tube. With a fairly high voltage already placed on the two electrodes in the tube, the electrons provide a brief conductor between the electrodes, and, consequently, a surge of electricity flows through the tube. This surge can then be recorded on a counting device or sent to an audio device that produces an audible click. Each time a Geiger-Müller counter clicks or tallies another count on its recorder, a particle of sufficient energy has entered the tube and made an ionized region.

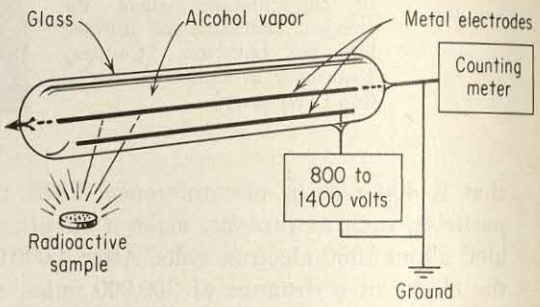


Fig. 25-11. The Geiger-Müller counter.

The ionized region caused by a radioactive particle may also be detected in a very different manner. We have seen previously that in an atmosphere that is supersaturated with water vapor, condensation to liquid water occurs on dust particles, forming clouds or rain. C. T. R. Wilson of England, in 1896, discovered that ionized gas molecules may serve as nuclei for the condensation of water droplets. Here is a striking example of the consequences of scientific curiosity and observation. Wilson, while yet a young student, was concerned about cloud formations, and, upon examining the effect of X rays upon such formations, he noted that fog droplets form in dust-free air when it is excited with X rays. This observation led to the development of the famous *Wilson cloud chamber.*

In a cloud chamber, a region of supersaturated vapor may be produced by either pressure changes or temperature changes comparable to those occurring naturally in weather patterns. That is, a saturated atmosphere becomes supersaturated with vapor when its pressure or its temperature is decreased. One method of producing a continuously supersaturated region is shown in Fig. 25-12. Alcohol evaporates from the saturated pad at the top of the chamber and condenses on the cold plate at the bottom. There is a continuous migration of the vapor from top to bottom, and at a certain level, the air in the chamber is saturated. Just below this level, where the air has a tendency to be supersaturated, high-energy particles or radiation, such as X rays, gamma rays, and cosmic rays, produce ionization trails that immediately form fog trails that are readily visible. Figure 25-13 is a photograph of fog trails produced by the alpha,

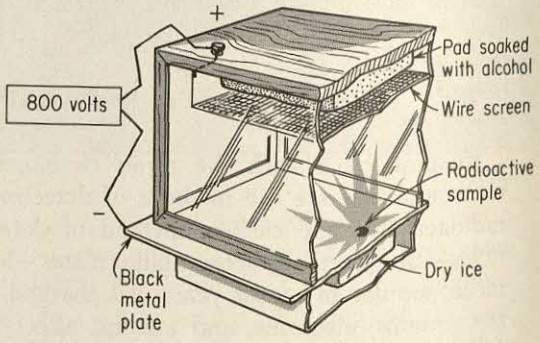


Fig. 25-12. A cloud chamber.

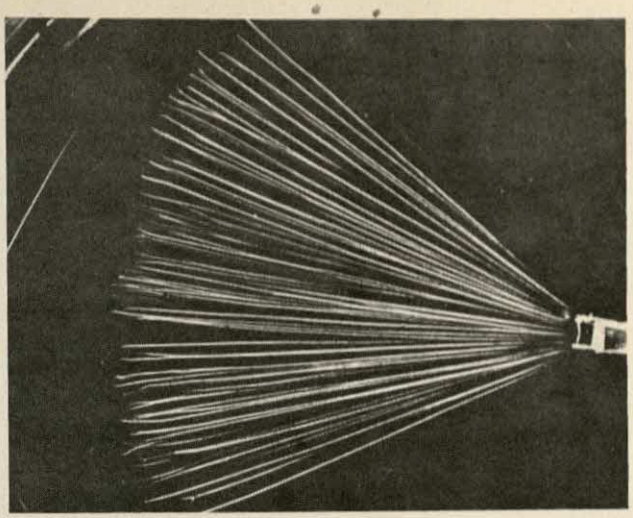


Fig. 25-13. Cloud chamber tracks produced by alpha particles emitted by a trace of polonium. (From Gentner, Maier-Leibnetz, and Bothe, *An Atlas of Typical Expansion Chamber Photographs.* London: Pergamon Press, 1954)

beta, and gamma rays evolved from a very small quantity of a radioactive substance.

The cloud chamber is a tool by which several important properties of the atomic nucleus have been discovered. For example, the existence of mesons was first proved through a study of photographs of cloud chamber tracks. Figure 25-1 is a photograph of fog trails that were produced by cosmic rays. These rays are extremely high-energy particles which, upon entering our atmosphere, may produce high-energy gamma rays, and we shall have more to say about them in Chapter 29.

A very sensitive method of measuring radioactivity is through the observation of the luminescence of certain substances when they are struck by alpha particles. If a thin layer of zinc sulfide is placed near a radioactive substance, a large number of individual flashes may be seen on the zinc sulfide layer. A simple method of viewing these flashes is shown in Fig. 25-14. Each flash represents the collision of an alpha particle with the zinc sulfide layer. Such a device is called a *scintillometer,* and in recent years it has been radically improved by using substances more sensitive than zinc sulfide and by detecting the flashes by electronic means. Such an instrument has become quite popular with the prospector who wishes to survey the radioactivity of certain areas from an airplane or automobile as well as on foot. We may observe directly the phenomenon of scintillation by examining the luminous numerals of a watch or clock with a magnifying glass in a darkened room. The glow of the numerals will appear as tiny flashes of light. The luminous paint contains an alpha emitter and zinc sulfide; each tiny flash of light is the result of the crash of an alpha particle with the zinc sulfide.

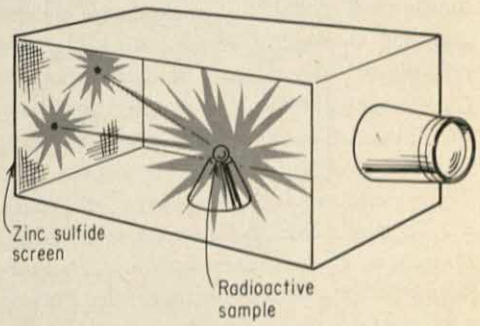


Fig. 25-14. A radioactive sample produces scintillations on a zinc sulfide screen.

25-7 Radioactivity in Tracer Studies

The availability of artificial isotopes, both stable and radioactive, has opened up an entirely new area of scientific

study. It has been predicted that eventually the greatest contribution of the atomic energy program will be through the use of these isotopes. Let us consider a few examples of this kind of experimental technique.

Through transmutation reactions, ordinary carbon (atomic mass 12) can be converted to carbon-14, which is radioactive because it is a beta emitter. Since, for all practical purposes, C-12 and C-14 are chemically identical, an organic compound may contain C-14 in place of ordinary C-12 anywhere in its molecule. The presence of the C-14 may be detected by any of the ordinary instruments, such as a Geiger-Müller counter. The radioactivity of C-14 serves as a "tag" or "label" by which we may follow it through any number of chemical processes, and we call such an isotope a radioactive *tracer element*.

Suppose we make some acetic acid, CH_3COOH, in which one of the carbon atoms is radioactive C-14 and the other is ordinary C-12 ($C^{14}H_3C^{12}OOH$). Then let us feed this substance to an organism that digests the acetic acid (acetic acid is the active component in ordinary vinegar). Among the digestion products are carbon dioxide and another substance. Since the carbon dioxide is not radioactive and the other substance is, we may readily conclude that the carbon dioxide is derived from the carboxyl group of the acetic acid. It is impossible to make this distinction by chemical means alone, and, therefore, the use of a tracer element is the only means by which this problem could be solved.

In Chapter 18 we saw how carbohydrates, fats, and proteins are made in the photosynthesis process. In the crucial studies of this process, labeled carbon (C-14) and labeled phosphorus (P-32) were used as plant foods. The pathways of the carbon and the phosphorus were followed by detecting their radioactivities. Again, a plant fed with radioactive phosphorus accumulates the phosphorus in various parts of the stem and leaves. By placing the plant on a photographic film, the radioactive rays from the phosphorus darken the film and give a direct record, known as a *radioautogram*, of the presence of the phosphorus (Fig. 25-15).

Fig. 25-15. A radioautogram of a bean plant fed with radioactive Calcium^{45}. Note the concentration of Calcium in the roots. (Courtesy of Orlin Biddulph, Washington State University)

In recent years the amount of tritium, a heavy isotope of hydrogen, in rainwater has been measured, and, since this isotope was originally formed from atmospheric water vapor, such a measurement gives us an indication of the age of the water. Cosmic rays probably account for the conversion of ordinary hydrogen to tritium, which has a half-life of 12.5 years. The age of river waters, the age of glaciers, and even the age of wines have been successfully studied by this technique.

Industrial applications of tracer elements are common. The mechanism of friction in engine bearings, the examination of welds and castings, and the measurements of the thicknesses of paper, plastics, and sheet metal have been made. In pumping various oils through a pipeline, the end of one batch and the beginning of another is marked with a

short-lived radioactive isotope. At the receiving station, when the operator notes a response on a Geiger-Müller counter, he turns the valves, thus separating the two batches of oil that came one after the other through the same pipeline.

25-8 The Measurement of Time by Radioactivity

In Chapter 2 we saw briefly that the age of the earth's crust has been determined by radioactivity. Now we are in a position to examine in more detail how these determinations are made. They are all based on the primary idea that radioactive decay occurs independently of any outside influence, such as temperature, pressure, and catalysts. Apparently the nuclei of the naturally radioactive elements are as unstable now as they ever were. That is, the intensity of radioactive emission by a particular isotope, as indicated by its half-life period, is constant. Here we have one of the most interesting applications of radioactivity—a long-range clock that measures the time of happenings in the past.

The recent past can be dated effectively by a method based on the carbon cycle in nature. We know from ordinary experience that carbon is present in the air as carbon dioxide. Plants use up this carbon dioxide, forming carbohydrates and other plant products that are then consumed. The carbon eventually returns to the atmosphere as carbon dioxide. However, some of the carbon dioxide in the air is composed of the carbon-14 that is continuously being made from atmospheric nitrogen by the transmutation:

$$ {}_{7}N^{14} + {}_{0}n^{1} \rightarrow {}_{6}C^{14} + {}_{1}H^{1} $$

Neutrons produced by cosmic rays change nitrogen to carbon and protons. Carbon-14 has a half-life of about 5760 years.

As a result of the continuous supply of atmospheric carbon dioxide containing carbon-14, all plants show a certain amount of radioactivity, which, though small, can be accurately determined. For example, each gram of carbon as it occurs in new wood gives off 12.5 counts of radioactivity per minute. If a piece of wood were preserved for 5760 years, it would give off only 6.25 counts per minute per gram of carbon. Thus, any sample of carbonaceous material that was once a part of the carbon cycle may be examined for radioactivity. As long as a plant is living it maintains a constant ratio of C-12 to C-14 in its structure, but when the organism dies, the intake of carbon ceases. Since the C-14 continues to undergo decay, the intensity of its radioactivity gives a direct indication of its age. The half-life characteristic of carbon-14 and the part it plays in the carbon cycle permit us to measure backward in time to possibly 40,000 years.

The carbon dating method, perfected by W. F. *Libby* at

the University of Chicago in 1946, has proved to be amazingly accurate. For instance, a specimen of wood obtained from an Egyptian tomb was estimated by archaeologists to be 4600 years old, and the radioactivity of the carbon-14 in this sample of wood showed it to be 4500 years old. The carbon dating method is accurate to within less than 50 years. Thus we have a dating method that appears to be dependably accurate and essentially independent of contemporary history. No matter what the preatomic-age history of the object, if it contains carbon, its age can be determined when proper corrections are applied. Recent atmospheric testing of nuclear weapons and the tremendous addition of carbon dioxide to the atmosphere during the industrial age complicate the job, but for the study of archaeological and geologic objects this is one of man's most accurate and useful tools.

The age of the earth has captivated the interest of mankind throughout the years. Before the discovery of radioactivity, scientific estimates ranged from 20 to 90 million years, based on studies of the composition of the ocean, the rate of the earth's cooling, and the separation of the moon from the earth. These studies, however, suffer from an important error, because they assume that a given process takes place at a constant rate. Although this assumption is necessary in order to reach a conclusion, it is certain that the rate of erosion of the continents has changed in time, and that while the earth was cooling, radioactive elements were adding heat to the earth. Since the total change in any of these processes is not known, the figures obtained by these methods may fall far short of the actual time interval under study.

More recently, the age of the earth's crust has been determined by at least eight independent methods in which the rate of change in the process is quite well known, and, surprisingly enough, they are in good agreement with one another. Even though they differ by as much as a factor of 2, this is not a serious discrepancy considering the great intervals of time involved and the widely different methods of measurement employed. Table 25-2 lists a number of the radioactive decay systems that have been used for these measurements.

Measurements using these systems have given the age of

TABLE 25-2

SEVERAL DECAY SYSTEMS USED TO DETERMINE GEOLOGIC AGES

Parent element	*Decays by giving off*	*Derived element*
U^{238}	$8\alpha + 6\beta$	Pb^{206}
U^{235}	$7\alpha + 4\beta$	Pb^{207}
Th^{232}	$6\alpha + 4\beta$	Pb^{208}
Rb^{87}	β	Sr^{87}
K^{40}	Electron capture	Ar^{40}

the earth's oldest rocks to be between 3 and 3.5 billion years. From this and much additional evidence, scientists presently set the age of the earth at about 4.5 billion years. This billion-year gap may have two explanations: rocks on earth fail to preserve their age record beyond 3.5 billion years, or granitic rocks were not made and stabilized within the first billion years. This whole area of study is a very active field of research and new evidence that we may obtain from the exploration of the moon may shed important new light on the problem. But let us now look in more detail at some of these geologic clocks.

Only those isotopes with relatively long half-lives are suitable for determining geological ages. From the half-lives of those isotopes that still exist in the earth's crust in appreciable quantities, we may estimate the upper limit of the earth's age. For example, we can tell from the half-life periods of uranium-235 (7.1×10^8 years) and potassium-40 (1.27×10^9 years) that the earth's crust is probably no older than 10^{10} years. If it were much older than this, these two isotopes would no longer be found in the earth's crust.

The uranium radioactive series, ending with lead-206 (Fig. 25-2), provides the essentials of the *lead clock*. A uranium mineral is analyzed carefully for the amount of lead-206 that it contains. Since lead-206 can be derived only from uranium-238, a knowledge of the amount of lead-206 will tell us how long the mineral has been in existence. The relationship that gives the age of the mineral is:

$$\text{age} = \frac{\text{amount of uranium-238}}{\text{amount of lead 206}} \times 7600 \text{ million years}$$

The factor 7600 million years is derived from a knowledge of the half-life periods of all the isotopes involved in the decay series. If a particular uranium mineral has a lead/uranium ratio of 0.132, then the age of this mineral is 0.132×7600 million years, or *one billion* years. One of the oldest minerals of this type was found in Southern Rhodesia; it has an age of 2.7 billion years.

Similar to the lead clock is the *helium clock*. In the series of decay reactions from uranium to lead, eight alpha particles are produced. Soon after an alpha particle is emitted from the nucleus of an atom, it picks up two electrons and forms a helium atom. Since alpha rays have only a short range in dense material such as rocks and minerals, the resulting helium atoms are entrapped in these solids. It is possible to measure the amount of helium in radioactive samples, and from the ratio of the amount of helium to the amount of uranium, the age of the mineral can be estimated. The equation relating these factors for the helium clock is:

$$\text{age in millions of years} = \frac{\text{cm}^3 \text{ of He in 100-g sample}}{\text{g of U in 100-g sample}} \times 8.8$$

Since a gram of pure uranium gives off only 1×10^{-5} cm^3 of He per year, for the helium clock to be accurate, careful

analysis of the samples is required. Even under these difficulties, scientists have discovered that the helium clock gives time measurements that agree well with those found by the lead clock.

In recent years, the age of the earth's crust has been measured by what is believed to be an even more dependable relationship—the relative abundance of isotopes. These measurements are even more accurate for determining the age of the earth itself, because they do not depend on the composition of an undisturbed sample of mineral. In nature there are four stable isotopes of lead, having the masses 204, 206, 207, and 208. Only lead-204 is of nonradioactive origin; the others have been produced as end products in a radioactive decay series. Lead minerals have been carefully analyzed by the mass spectrograph for isotopic content. It is apparent that when a lead mineral becomes part of the earth's solid crust its composition is no longer changed. However, since radioactivity adds additional lead-206, 207, and 208 to the earth's crust in certain places, a comparison of the isotopic composition of "primeval" lead to "radioactive" lead gives a measure of the age of the earth's crust. Similar measurements have indicated that the earth and meteoritic material had similar origins, but we shall make a more thorough examination of this idea in Chapter 27.

Finally, there are *pleochroic halos*, or colored rings, in certain minerals such as biotite and fluorspar. It is well known that high-energy radiation causes coloration in many materials. For instance, ordinary glass becomes purple when exposed to sunlight for extended periods of time, and a person who works with radioactivity and is exposed to high-energy radiations wears a small disk of material that changes color in a similar way. In rocks and minerals that contain a naturally radioactive isotope, the colored rings called pleochroic halos are formed by this coloration effect. The size and color of these rings, compared to rings produced artificially, give a measure of the age of the mineral. Although it is an interesting phenomenon, this method of measuring time is not as accurate as others, and it is no longer widely used.

25-9 Summary

Radioactivity has provided a key that has enabled us to unlock the inner structure of the atom. The characteristics of the alpha, beta, and gamma rays emitted by radioactive processes indicate that the particles within the nucleus—protons, neutrons, neutrinos, mesons, and so forth—occupy energy levels comparable to the energy levels occupied by electrons in the outer parts of the atom. The energies of the nuclear particles are, however, at least a million times greater than the energy of the extranuclear electrons.

Radioactivity is the spontaneous decay of the atomic nu-

cleus in which alpha, beta, and gamma rays are emitted. When an atom emits an alpha particle, its mass decreases by four units and its atomic number decreases by two. A beta particle, emitted from the nucleus, increases the atomic number by one.

The stability of the atomic nucleus is determined by its binding energy. This can be found by determining the difference between the total masses of the particles making up the nucleus and the mass of the nucleus itself.

Transmutations of one element into another, and the change of the isotope of one element into another isotope of the same element can be achieved by adding, at sufficient energy levels, either alpha particles, protons, neutrons, or other nuclear particles, to the atom's nucleus.

The energies needed for transmutation reactions are obtained by accelerators such as the cyclotron, bevatron, and linear accelerator. These devices, though complex in construction, are essentially an application of the basic principles of electrostatics, electricity, and magnetism.

High-energy particles are detected by ionization chambers—the Geiger-Müller tube and the cloud chamber—and by luminescent materials such as zinc sulfide.

Radioactivity provides a powerful tool for the study of a host of scientific and technological problems through what is termed the tracer technique. Radioactivity is also widely used in the study of extended time intervals such as the age of carboniferous deposits and the age of the earth.

EXERCISES

A. KEY TERMS TO REMEMBER

bevatron
binding energy
carbon dating
cyclotron
Geiger-Müller counter
half-life period
helium clock
lead clock
meson
mev
nuclear energy barrier
pleochroic halos
scintillation
tracer element
transmutation
Wilson cloud chamber

B. QUESTIONS ABOUT RADIOACTIVITY

1. Make a list of the particles that may be added to or removed from the nucleus of an atom.

2. List as many characteristics as you can that are common to the radioactively stable elements in nature.

3. Explain why a change in the number of protons in the nucleus of an atom changes the atom to a different element.

4. What are the primary purposes of the cyclotron, bevatron, and similar devices?

5. Distinguish between natural radioactivity and artificial radioactivity.

6. Of the methods of measuring radioactivity, which ones do

you believe are capable of detecting all the particles that are emitted by a particular radioactive sample?

7. What factors may have caused the rate of salt accumulation in the oceans to vary through the ages?

8. Tell how radioactive isotopes may be used in medical studies.

9. What assumptions are made in dating the earth by the uranium clock method?

10. Examine a luminous dial of a watch or clock in a darkened room, using a magnifying glass. Explain the cause of the tiny flashes of light that you observe.

11. What assumptions are made in dating an archaeological find by the carbon-14 method?

C. PROBLEMS ABOUT RADIOACTIVITY

1. On the basis of known masses for the neutron and the hydrogen atom, calculate the mass defect of the nitrogen atom, atomic mass 14, atomic number 7. $H^1 = 1.0081$, ${}_0n^1 = 1.0090$, N = 14.0080. Ans. 0.1117.

2. Using the Einstein equation, calculate the energy released by the formation of a gram-atomic weight of carbon from 6 hydrogen atoms and 6 neutrons. ${}_6C^{12} = 12.0000$.

3. Using a material like plaster of paris, construct a model of the energy barrier shown in Fig. 25-7. Using marbles for the positively charged particles, determine experimentally the probability of successfully rolling a marble over the energy barrier and into the energy well.

4. Iodine-124 is radioactive and decays with a half-life of 12 days. If we ate 0.1 gram of iodine-124 and all of it was retained within the body, how much I-124 would be in the body after 24 days? Ans. 0.025 g.

Twenty–Six

Nuclear Energy

The discovery of practical methods of unleashing the energy of the atom's nucleus has thrust upon mankind political, military, and economic as well as scientific problems undreamed of prior to 1939. Nuclear energy has become a decisive factor in national and international relations, and it may eventually be the major source of cheap power for the expanding needs of our modern civilization. Already, several nuclear electrical generating plants are in operation, and the possible use of nuclear energy for large construction tasks is indeed bright. Few discoveries have seriously altered the course of history more than nuclear energy. It brought a major war to an abrupt close, and it now provides the chief deterrent to other major conflicts.

Nuclear energy may appear to us as an untamed lion among lambs, and we may be baffled by its immensity and fearful of its possibilities. Yet, as with such inventions as gunpowder, the airplane, and radar, it may serve a constructive as well as a destructive purpose. By-products of nuclear energy have provided the means to study nature in a manner not otherwise possible. The mechanism of photosynthesis, the age of the earth and solar system, and the origin of the universe are problems that can now be studied more adequately through tools derived from nuclear energy changes.

Let us examine the magnitude of nuclear energy in terms of some of our common experiences. Ordinary physical changes usually involve intermolecular forces, and, for example, the melting of ice requires only that we break or disrupt the forces between molecules. This requires about 80 calories of energy per gram of ice. If we wish to tear apart the water molecules themselves, over 50 times as much en-

ergy is required. In such a chemical process the outermost parts of the atoms in the water molecules are changed. We frequently resort to these chemical reactions for energy sources. For example, a pound of coal burned to carbon dioxide, releases slightly over 12,000 Btu, or 4 kilowatt-hours of energy. Whether we burn coal, gasoline, or rocket fuel, we are tapping the energy that is stored up in the outer electrons of the atoms.

In even the most energetic chemical processes, nothing whatever happens to change the nucleus. The inner parts of the atom remain unaffected. When the components of the nucleus are altered, quantities of energy are involved that are of almost unbelievable magnitude. In the simplest of nuclear changes, radioactivity, the energy that is released is enormous, some millions of times greater than that released in chemical reactions. These huge quantities of energy arise from the binding energies that hold the neutrons and protons together within the nucleus, and they are well beyond the limits of our ordinary experience. The alpha particles alone that are ejected from one pound of radium release over 800,000 kilowatts of energy. In a vacuum, where no gas molecules slow them down, the velocity of alpha particles from radium is 3.6 million miles per hour! Table 26-1 makes

TABLE 26-1 ENERGY RELEASED IN VARIOUS PROCESSES

Reaction	*Energy released in electron-volts per atom or molecule of material*
Melting of ice	0.02
Burning of coal	3.5
Thermite reaction	6.7
Explosion of TNT	16
Emission of alpha particles from uranium-238	$.2 \times 10^6$
Fission of uranium-235	1.98×10^8

a comparison of these energies to those derived from several common physical and chemical processes.

26-1 The Source of Nuclear Energy

There is a striking difference between nuclear energy and ordinary chemical energy. Chemical energy is derived from electron changes in the outer part of the atom, whereas nuclear energy is concerned with changes within the tiny nucleus. Let us examine what is known about the nucleus in order to find the source of the huge amounts of energy that may be obtained from it.

The forces that exist between two particles vary with the distance between them. At great distances, these forces may

be gravitational, electrical, or magnetic, depending upon the nature of the particles, and they vary inversely as the square of the distance between the particles. But at extremely short distances, a new kind of force is observed that varies inversely with a much higher power than the square of the distance. For instance, the force of repulsion between two protons increases as the square of the distance between them decreases (Coulomb's law) and at 10^{-12} cm, this repulsive force is equivalent to about 6 tons. Within the nucleus at distances less than 10^{-12} cm, however, a much different kind of force takes over—a nuclear force of attraction equivalent to about 240 tons. This strange change in affairs as we approach the small dimensions of the atom's nucleus is related to the source of nuclear energy and is responsible for holding together the positive and neutral particles within the tiny nucleus of the atom. The forces acting between nuclear particles are more than a hundred million times stronger than the forces that hold an electron in its position in the outer part of the atom. This difference may be compared to that between the ring of an alarm clock on a bedside table and the ring of one over 50 miles away.

In the spontaneous decay of radioactive nuclei, we saw that alpha particles and beta particles are emitted from the nucleus. The alpha particle is composed of two protons and two neutrons. At extremely short ranges, these four particles hold each other together with a force much greater than is the force of repulsion between the two positively charged protons. Apparently the alpha particle, even though bound to the nucleus by other exchange forces, has forces of its own that hold it together when it is ejected from the nucleus. When it escapes from the nucleus, we may think of it as rolling down hill to a much lower energy level (Fig. 26-1). Consequently, it releases energy equivalent to the energy of the forces that are in the nucleus. It is interesting to note here that the amount of energy released by an alpha particle is determined by the particular element from which it comes. Fog tracks of alpha particles released from a single isotope have identical lengths and widths, showing that they have identical energies (Fig. 25-13).

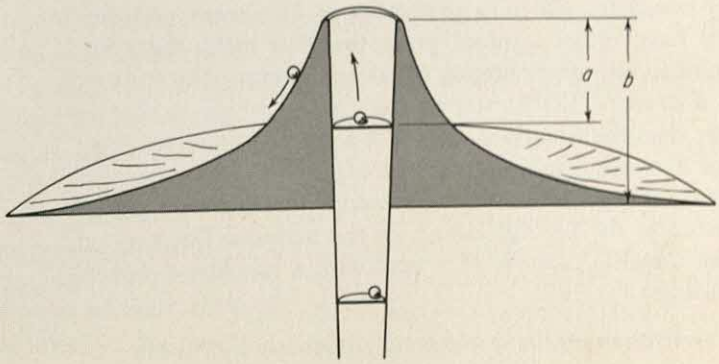


Fig. 26-1. The energy trough of a nucleus that is about to emit an alpha particle.

In recent years, considerable attention has been given to the study of other kinds of particles that may be in the nucleus of the atom. If enough outside energy is provided to lift a particle out of the energy trough of the nucleus, the energy mass and charge of this particle may be examined in the cloud chamber. Sufficient energy for these studies is obtained from two sources—cosmic rays and an artificial accelerator such as a cyclotron. When the nucleus of an ordinary atom is struck by a cosmic ray or by one of the high energy particles from a cyclotron, some of the particles within the nucleus are given enough energy to overcome the binding forces that hold them in the nucleus. The fragments that are chipped off are usually very short-lived, some of them lasting only 10^{-24} seconds! However, they can be detected and studied, and some indication of the structure of the nucleus can be gained from them. Along with the mesons (Table 25-1) there are upward of two dozen different particles that have been detected in the atom's nucleus. The problem of fitting this bewildering array of new particles into a reasonable picture is one of the most exciting puzzles that faces the physical scientist. When these particles are finally arranged in order, we will no doubt be closer to a true picture of the source of nuclear energy. In the meantime, we must be satisfied with only a partial answer to the question of the source of nuclear energy. But we can be reasonably certain that nuclear energy is associated with two unusual dimensions—extremely short distances and extremely short time intervals. That is, nuclear energy, which is about 10^{40} times greater than gravitational energy, is associated with tiny particles that are only 10^{-13} cm from each other, and the time intervals involved in nuclear energy changes are very short. We might say that these two forms of energy occupy opposite ends of the scales of time and distance that exist in the universe.

26-2 The Neutron

Before we describe the means by which we may tap the reservoir of nuclear energy, it may be well to consider in more detail the neutron. Neutrons are involved in many nuclear reactions, either as projectiles or as ejected particles. They have rather unusual properties that make them important in any consideration of nuclear energy. The neutron has a mass of 1.0090 atomic mass units, which is slightly larger than the combined masses of a proton and an electron. Since a neutron may be considered as a combination of a proton and an electron, this difference in mass must be attributed to the binding energy in the neutron. The neutron is not a stable particle. It decays with a half-life of twelve minutes:

$$\text{neutron} \rightarrow \text{proton} + \text{electron} + \text{neutrino} + 0.78 \text{ mev}$$

For this reason we should never expect to find an accumulation of free neutrons. Furthermore, the neutron has no charge, and if it approaches the nucleus of an atom, it stands a good chance of falling into the energy trough of the nucleus (Fig. 26-2) and being captured.

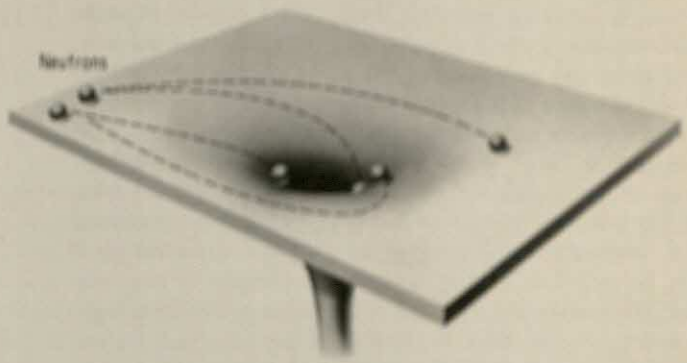


Fig. 26-2. To a neutron, the energy trough of the nucleus appears to be a sink.

Whether a neutron is picked up by a particular nucleus depends on the energy of the neutron, and we find that its energy may vary between wide limits, depending on its source. When neutrons are evolved from the nucleus of a carbon atom, they have 5.5 million electron volts of energy:

$$_4Be^9 + {}_2He^4 \rightarrow {}_6C^{13} + {}_0n^1 + 5.5\ \text{mev}$$

Neutrons may be obtained from several convenient transmutation reactions similar to this alpha bombardment of beryllium. The alpha particles from one gram of radium, when striking beryllium, will yield 10 to 15 million neutrons per second. Neutrons may also be produced by reactions induced by the high-speed accelerators. However, it is not possible to accelerate neutrons in the same way that other atomic and nuclear particles are accelerated, simply because they do not carry a charge.

On the other hand, it is sometimes desirable to slow down neutrons in order to bring about certain reactions involving them. The ideal material for slowing them down should provide a medium through which the neutrons can pass and not be captured. Such a material is called a *moderator*, and as the neutrons pass through it they may collide with a number of its atoms. After the neutrons have collided with enough atoms of the moderator, they are slowed down. When their kinetic energy has been reduced to that of ordinary atoms and molecules, the neutrons are called *thermal neutrons*, and although they may have only a fraction of an electron volt of energy, they are still capable of entering into nuclear reactions. This is indeed a remarkable property of the neutron, for when all other particles, such as protons, electrons, and alpha particles are similarly slowed down they no longer can react with atomic nuclei because they have

insufficient energy to penetrate the energy barrier of the nucleus.

The moderators best suited for the job of slowing down neutrons are the lighter elements. The hydrogen atom is ideal because its mass is approximately the same as the neutron. Each collision with a hydrogen atom results in the loss of almost 60 per cent of a neutron's energy, a loss that we might expect by applying Newton's third law of motion. In each collision some of the energy of the neutron is imparted to a hydrogen atom, and after only 18 collisions, the neutron will be slowed down to a thermal energy level. When a neutron is shot through carbon atoms, as many as 114 collisions are necessary to "cool" the neutron through a similar energy range. Even though carbon is not as efficient as hydrogen for this purpose, its ease of handling makes it a commonly used moderator for slowing down neutrons, a process that may be called a "cooling" process. In Chapter 20 we saw that temperature is a measure of the average kinetic energy of molecules, and since the energy of neutrons may be reduced by slowing down the neutrons, the process by which it is done is one of cooling.

Neutrons are almost completely stopped by some substances, which are said to have nuclei with large cross-sectional areas. Cadmium is one of these substances. When neutrons strike a sheet of cadmium metal, they are completely stopped, and the slower neutrons are readily "soaked up" by the cadmium nuclei. We shall see in a moment the important roles played by carbon and cadmium in neutron reactions.

26-3 Nuclear Fission

Since the discovery of artificial transmutation, nuclear scientists have tried to bring about transmutations of many elements. With the periodic table (see p. 254) as a guide, elements from the lightest to the heaviest were bombarded with protons, deutrons, alpha particles, and neutrons. Consequently, hundreds of new isotopes of the known elements were produced by man for the first time. Important nuclear changes can be achieved through bombardment with neutrons. When the nucleus of a target atom is hit by a neutron, the resulting nucleus is usually unstable; that is, the new, heavier nucleus undergoes emission of an alpha or beta particle in order to reach stability. This may be shown diagrammatically in Fig. 26-3, a process that has been observed in many nuclear changes.

In 1934, the Italian physicist, Enrico Fermi, sought to bombard uranium atoms with neutrons. His purpose was to create atoms heavier than uranium with the idea that uranium atoms may have evolved from heavier elements that had died out in the distant past. According to the above scheme, a uranium-238 nucleus, when it receives a neutron, should

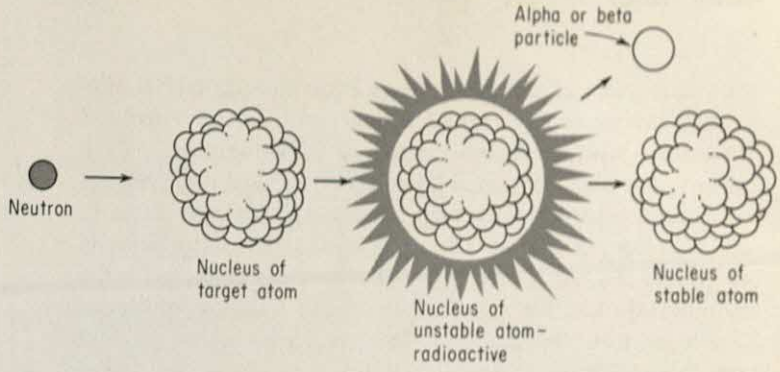


Fig. 26-3. A typical chain of events following the bombardment of an atom by a neutron.

go to uranium-239, then lose a beta particle and form a new element with one more proton in the nucleus than uranium has. This would be element number 93. But Fermi did not get the results he had hoped for. In place of a heavier element, several lighter elements appeared among the products of his nuclear reaction, but the proper analysis of these products was difficult because only extremely small quantities of them were produced.

Finally, Hahn and Strassman in Germany were able to isolate a small quantity of barium in the reaction products. When their report was made in January, 1939, there was a veritable explosion in the scientific world. Never before had anything greater than an alpha particle been expelled from an atomic nucleus. If the heavier uranium atom had ejected a barium atom, this implied that the uranium nucleus had literally split in two. This process of cleavage or breaking of the nucleus into much smaller pieces is now called *nuclear fission.* Within less than a month after this revolutionary report was made, it was definitely verified by no less than half a dozen of the most capable scientists of the time.

STATE INSTITUTE OF EDUCATION
Govt. of West Bengal

With the discovery that the uranium atom, when struck by a neutron, undergoes fission, attention was turned to the composition of the products of this peculiar reaction. It was soon discovered that among the products were several new neutrons, released along with the other fission products (Fig. 26-4). This posed the possibility that the neutrons that are given off may be used by new uranium atoms, thus setting up a chain reaction of nuclear fission. It is now common knowledge that such a chain reaction was soon developed.

The fission of uranium-235 is much simpler than the fission of uranium-238. That is, slow neutrons bring about the fission of the lighter isotope, whereas it requires high-energy neutrons to cause the fission of uranium-238 (as well as that of many other fissionable elements). Hence, the fission of uranium-235 has been studied extensively. A typical example of a nuclear fission reaction is:

$$_{92}U^{235} + {}_{0}n^{1} \rightarrow {}_{36}Kr^{84} + {}_{56}Ba^{138} + 3{}_{0}n^{1} + \gamma \text{ rays} + \text{energy}$$

The total energy released when *one* uranium-235 atom undergoes fission is approximately 200 million electron volts.

For each *gram* of uranium, this is equivalent to 60,000 kilowatt-hours of energy. About 80 per cent of this energy is associated with the kinetic energy of the fragments, which means that the two small atoms, barium and krypton, fly apart at a phenomenally high velocity, possibly as much as 25,000 miles per second. The neutrons have 5 million electron volts, and the remainder of the energy is associated with the gamma rays and the radioactivity of the fragment elements.

The neutrons that are evolved in the above fission reaction are too energetic to be picked up immediately by other uranium atoms. However, if the piece of uranium is large enough to slow down the 5 mev neutrons to thermal neutrons, then a chain reaction may proceed. From this fact we can see that there is a critical size for sustaining a chain reaction in uranium-235, and this size, probably no more than several pounds, permits the construction of military weapons based on this reaction.

The fission reaction has given indisputable support to the relationship between mass and energy. In the last chapter we examined the equation that was developed by Einstein through purely theoretical considerations (Equation 25.1). If we start out with 1000 grams of uranium-235, a complete account of the mass and energy that are involved in the fission process is as follows:

Original mass of U^{235}		1000 g
Mass of fission products	989 g	
Mass of emitted neutrons	10.2 g	
	999.2	
Net mass converted to energy		0.8 g

In the equation $E = mc^2$, the energy equivalent to 0.8 g of mass is:

$$E = 0.0008 \text{ kg } (3 \times 10^8 \text{ m/sec})^2 \text{ joule}$$
$$= 7.2 \times 10^{13} \text{ joules}$$
$$= 2.7 \times 10^7 \text{ horsepower-hours}$$

The energy released by a kilogram of uranium provides enough energy to power an ocean liner for many weeks and a typical freight train (which requires an engine of 5000 horsepower) for a full year without refueling. If this energy were to be released suddenly, it would be equivalent to the explosive force of 20,000 tons of TNT, one of the most powerful chemical explosives.

In order to use the energy from the fission process over an extended period of time, the rate of the chain reaction must be controlled, and in order to control this rate, a major installation is required. The generation of power from nuclear energy is considered in the next section.

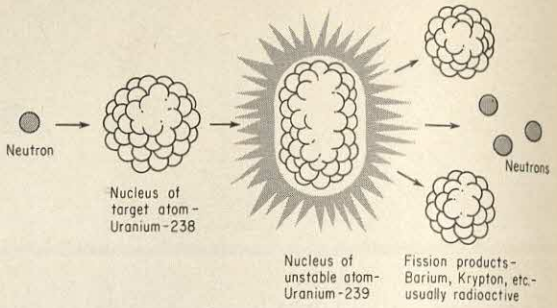


Fig. 26-4. The fission process of uranium-238.

26-4 The Utilization of Nuclear Energy

The elaborate efforts to produce more efficient chain reactions for military purposes have given us much information

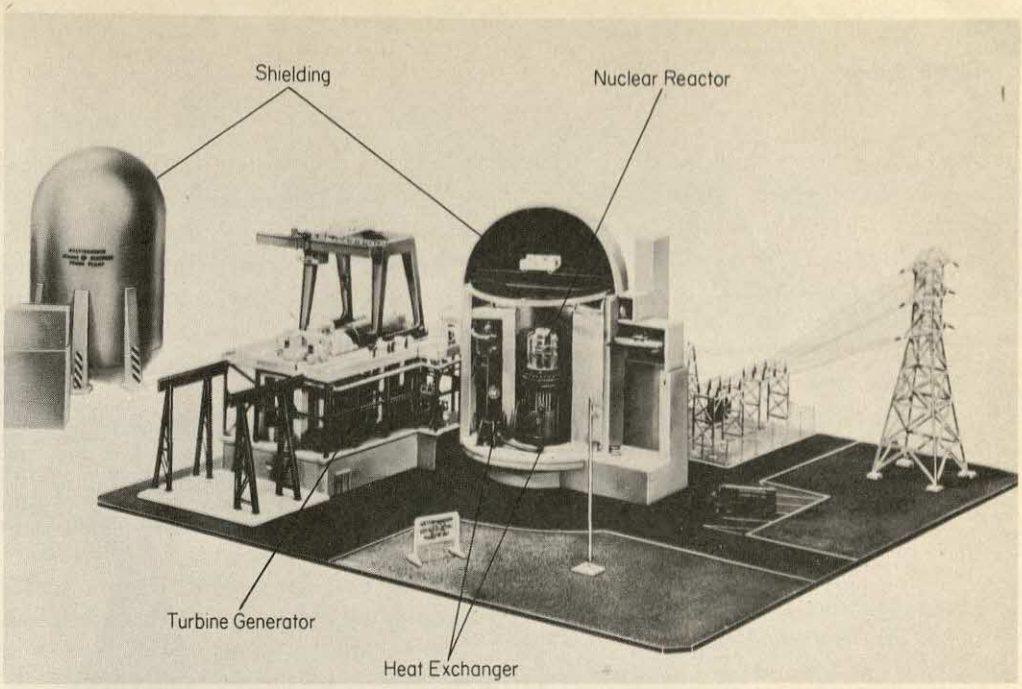


Fig. 26-5. A model showing the main components of a nuclear power generating plant. (Revell, Inc.)

regarding the peacetime use of nuclear energy. The earliest effort that successfully detected the chain reaction of nuclear fission was conducted in what is known as a *reactor*. Since this is essentially the process for conducting a controlled chain reaction, let us consider its general features. A conventional reactor is a stack of uranium rods embedded in graphite (Fig. 26-5). The graphite serves as a moderator, slowing down the neutrons to thermal levels. In order to control the chain reaction, movable cadmium rods are placed in the reactor. Since the cadmium absorbs neutrons, these rods serve as a governor for the process, for if neutrons are produced so rapidly that the chain reaction picks up speed, the cadmium rods can be moved into the reactor. The resulting removal of neutrons by the cadmium slows down the chain reaction.

As a chain reaction of nuclear fission occurs, there is a conversion of approximately 0.1 per cent of the fissionable material into energy. This energy is given off in the form of thermal or heat energy at extremely high temperatures. If suitable pipes are placed in the reactor, this heat energy may be continually removed from the reactor by a heat-transferring liquid and used to operate a steam turbine or a similar conventional heat engine. One type, used for military purposes, provides the power for new submarines and is proving quite successful (Fig. 26-6). The amount of uranium needed in a reactor may be quite small. For example, in the submarine engine, no more than five pounds of uranium-235 are necessary to supply the fuel for an extended run at sea.

The nuclear reactor enables us to use more of the potential nuclear energy in uranium than that which exists only in uranium-235. If we were to depend entirely on uranium-235 to supply us with nuclear energy, we would soon use up this particular isotope, for only a small fraction of natural uranium consists of this isotope. There is 140 times more uranium-238 than uranium-235 available. Fortunately, the

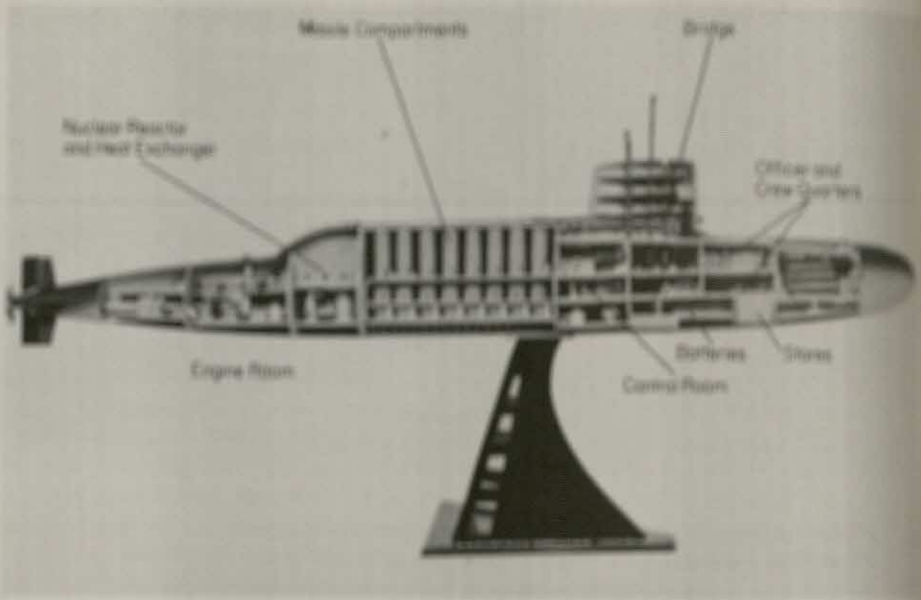


Fig. 26-6. A model of the nuclear powered Polaris submarine (George Washington class). (Revell, Inc.)

reactor provides an efficient means of using the energy of the uranium-238.

In the reactor, a stray neutron starts a process that goes something like this: a uranium-235 atom picks up a neutron, then fission occurs giving fission fragments and 2 or 3 neutrons (Fig. 26-7). One of these neutrons, after being slowed, may go into another uranium-235 atom and thus set up a chain reaction. But the other slow neutrons may be picked up by uranium-238, which is then converted into uranium-239. We have already seen that uranium-238 may undergo fission when struck by a fast neutron, but a slow neutron merely converts it into a radioactive isotope, uranium-239. The half-life of uranium-239 is only 23 minutes so that through beta decay it forms a new element, element number 93 (neptunium). Element 93 in turn loses another beta particle and forms element number 94 (plutonium), and in each of these steps energy is evolved as gamma rays. The final product of this chain of events, plutonium, is a new element that is readily fissionable. It is chemically different from the other elements in the reactor, hence it can be easily separated from the products by chemical processes. It can then be used again in a reactor where its fission reaction can be controlled, or it may be accumulated in a greater than critical size whereupon it undergoes a rapid chain reaction. Plutonium has apparently become a favorite material in producing military-type nuclear energy explosions. In general, the reactor is a versatile device for generating useful energy and for making the neutrons required for neutron induced transmutation reactions.

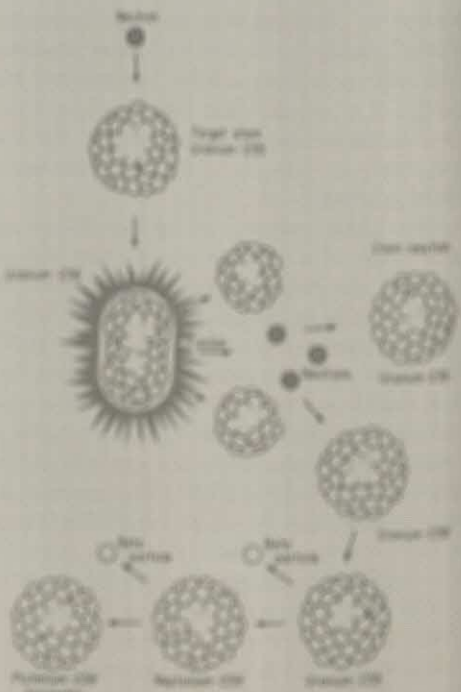


Fig. 26-7. The chain of reactions that occur among the uranium atoms in a reactor.

24-3 New Elements

The age-long desire of man to produce something new is finally realized in the breeder reactor. The process consists essentially of producing neutrons at suitable energy levels that permit the uranium atoms to become heavier through neutron capture. As neutrons are picked up, the resulting nucleus becomes unstable, and in order to achieve stability, the nucleus decays by beta emission and produces elements of higher atomic number. Several elements beyond uranium have been synthesized by this method.

With the exception of plutonium, no synthetic elements have been made that have found a practical use. However, the scientist in his search for more fundamental knowledge of the physical world has been keenly interested in the properties of these synthetic elements. They have enabled him to study many of the detailed relationships among the elements as they appear in the periodic table. It is often through such studies that important and far-reaching discoveries are made. It would be foolish for us to try to make a prediction of the industrial or commercial usefulness of americium, curium, berkelium, californium, or mendelevium, but as we look back among the many seemingly useless studies of the scientist, we discover that many of them have paid off in big dividends. At the time these studies were made, the scientist may not have been so concerned about the usefulness of his work as about merely adding to our store of knowledge.

For many years, unsuccessful attempts were made to find elements 41, 43, 85, and 87 in nature. After many fruitless years, the transmutation methods were developed by which these elements could be synthesized. When small samples of each of them were made, it became obvious why no one had found them in nature—they are radioactive elements that are short-lived. Since all natural radioactive series end with lead, there is no way in which a process of natural radioactivity could produce these lighter elements. Consequently, these elements have simply disappeared in nature.

In addition to several new elements, hundreds of new isotopes of the naturally occurring elements have been made. It does not concern us here to list many examples of these new isotopes, but one of the more useful products of artificial transmutation may be interesting at this point. When ordinary cobalt-59 is bombarded by neutrons in a reactor, the cobalt is converted into a radioactive isotope:

$$Co^{59} + n^1 \rightarrow Co^{60}$$

Cobalt-60 with a half-life of 5.3 years decays by the radioactive emission of a beta particle at 0.31 mev and of gamma rays at 1.17 and 1.33 mev. It is the gamma ray emission of cobalt-60 that is of therapeutic value in the treatment of certain types of cancer. Cobalt-60, fashioned into small rods, pellets, or needles, may be used in place of radium in cancer

treatment. Its cost is radically smaller, and it is more effective and more convenient to administer than radium.

26-6 Nuclear Fusion

Another source of nuclear energy, besides that obtainable from the heaviest elements, comes from the fusion of the lightest elements into heavier elements. In Fig. 25-5 we saw that the nuclear binding energy of elements of intermediate mass is much greater than the binding energy of the lightest and heaviest elements. We have already seen how we may obtain useful energy from the nucleus of a heavy atom such as uranium-235 by splitting up the atom into middle-weight atoms. A small fraction of the mass of the atom is converted into energy. But with the lightest elements, such as hydrogen, we must fuse together two or more atoms to make a heavier atom with the consequent conversion of mass into energy. Since these heavier atoms have more nuclear stability than does hydrogen, energy is released in the fusion reaction.

Hydrogen exists as three isotopes. The common hydrogen atom has a mass number of one. In each 5000 atoms of natural hydrogen there is one atom of mass number two, called deuterium. The third isotope, known as tritium, exists naturally to only an extremely small extent because it is radioactively unstable. It can be synthesized, however, in appreciable quantities. Between these isotopes, there are six possible combinations by which hydrogen can undergo fusion, forming atoms with heavier nuclei:

H^1	H^1
H^2	H^2
H^3	H^3

However, not all of these six reactions are feasible. In order to obtain useful energy from a fusion reaction, the reaction must take place at a reasonably rapid rate, and a sufficiently high temperature must be provided to trigger it.

Ordinary hydrogen atoms combine at an extremely slow rate, even when they are heated up to the triggering temperature. The reaction between two hydrogen-1 atoms is probably as follows:

$$ {}_1H^1 + {}_1H^1 \rightarrow {}_1H^2 + {}_1e^0 + 0.42\,\text{mev} $$

The time required for this reaction is 10^{11} years. It is believed that this extremely slow reaction occurs within the interior of the sun and accounts for the sun's predicted ability to give off energy for at least 30 billion years. Since it is difficult to maintain a reaction on earth at extremely high temperatures for even a second or two, we must use a fusion reaction that can be triggered in, say, a fraction of a second. Such a reaction is provided between a deuterium and a tritium atom:

$$_1H^2 + _1H^3 \rightarrow _2He^4 + _0n^1 + 17.6 \text{ mev}$$

This reaction occurs in the extremely short time of .0000012 seconds.

There is no critical size for a hydrogen fusion reaction. As small or as large an amount of an intimate mixture of deuterium and tritium may be ignited as we wish, provided that a high enough temperature is available to trigger the reaction. The uranium-235 fission reaction, if permitted to enter into a rapid chain reaction, produces sufficient energy to set off the hydrogen fusion. But once the fusion reaction starts, there is at present no way of controlling it. Thus, the hydrogen fusion reaction is of use for the time being only for military and scientific purposes.

The 17.6 mev of energy of the hydrogen fusion reaction may seem modest in comparison to the 200 mev of the uranium-235 fission reaction. However, the energy from the hydrogen reaction may add up to a significant amount if we take into account the differences in masses of the isotopes concerned. In the hydrogen reaction, the reacting masses are 2, 3, or 5, compared to 235 for the uranium reaction. Thus, on a pound-for-pound basis, the hydrogen fusion reaction is capable of providing almost four and one-half times as much energy as the uranium reaction. And, since only a high temperature is required to set off the fusion reaction, the fusion process is called a *thermonuclear* reaction.

Supposedly, since the fusion reaction is a chain reaction involving only the transfer of energy from one step of the chain to the next, it goes more nearly to completion than does the fission reaction. A considerable share of the uranium atoms probably never receive a neutron during a fission reaction and consequently they do not share in the chain reaction.

If we take all these factors into account, it becomes apparent that we can make a hydrogen fusion bomb as large as we wish. A fearsome and awful threat is posed if such a bomb were ever to be used. The Hiroshima bomb, equivalent to 20,000 tons of TNT, brought severe destruction to an area 1 mile in radius. A hydrogen bomb, supposedly 2000 times more powerful, would destroy everything in a circle 7 miles in radius, and the heat and shock effects would extend outward 20 miles or more. In other words, the city of Chicago and its suburbs could be wiped out in a single flash.

In addition to the immediate destruction of a nuclear energy bomb, there is the possibility of considerable after effects in the form of radioactive fallout. The large amount of free neutrons produced by the blast would be rapidly picked up by dust and gases, converting many otherwise stable elements into radioactive isotopes. Fortunately, the elements that are constituents of the soil and of ordinary building materials form only short-lived radioactive isotopes, and the radioactivity from them would soon fall to a harmless level.

The intense heat of the fusion reaction is believed to be

sufficient to initiate the fission of ordinary uranium-238 (Fig. 26-4), thus providing the means of building a superweapon. The first stage is a plutonium fission chain reaction, the temperature of which is great enough to start a fusion reaction in a mass of hydrogen. The temperature of this second stage is sufficient to start a third stage, the fission of uranium-238:

$$_{92}U^{238} + {}_{0}n^{1} \text{ (at very high energy)} \rightarrow \text{fission products} + \text{energy}$$

This three-stage process (called a fission-fusion-fission or 3-F bomb) would release energy equivalent to 20 million tons of TNT, and there is ample evidence to indicate that such bombs (Fig. 26-8) have been set off. The heavy fallout of radioactive by-products from the tests of so-called hydrogen bombs has been great enough to lead scientists to conclude that a large quantity of uranium had undergone fission. Neither the fission trigger nor the hydrogen fusion reactions would produce so large a quantity of radioactive fallout as has been observed. Needless to say, the potential dangers of the radioactive by-products of such a bomb are formidable hazards, and it behooves man to find other means of settling his international problems.

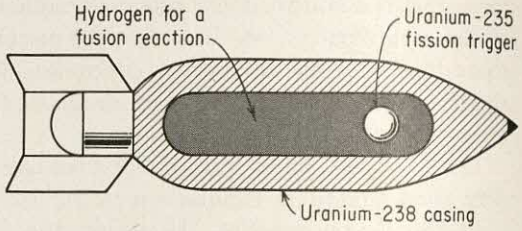


Fig. 26-8. A hypothetical fission-fusion-fission bomb.

26-7 Separation of Isotopes

We have already seen the importance of having a single isotope in relatively pure form, for it is impossible to have a chain fission reaction of a natural mixture of uranium isotopes. A mixture of hydrogen isotopes is of little interest in a hydrogen fusion reaction, and in preparing special isotopes for tracer use, a pure sample of an isotope is obviously preferred. Because the separation of isotopes may be a critical step in obtaining nuclear energy, let us examine a few of the methods used to separate them.

The first successful method and one that is still of prime importance is based on gaseous diffusion. We saw in the kinetic structure of gases that light molecules travel faster than do heavy molecules. Consequently, light molecules diffuse through a porous barrier more rapidly than do heavy molecules. This method is so sensitive that it may be used to separate the two gases $U^{235}F_6$ and $U^{238}F_6$, even though one of these is but one per cent heavier than the other. The installation required for this separation is elaborate because the separation cannot be completed in one simple step. The gases must be recycled through at least 4000 different porous barriers, and a large plant is required to house the pumps and various chambers needed in this operation. This process for separating uranium isotopes was placed in successful operation early in 1945, and the tremendous problems and difficulties in this operation warrant its being described as nothing less than an amazing feat.

We have already seen how the existence of isotopes was first proved—by means of a mass spectroscope. This instrument, based on the electromagnetic behavior of charged particles, or ions, may be used to separate isotopes. Since the actual quantity of isotopes that may be separated by this method is small, it is used primarily to separate isotopes that are used for special research work.

There are at least four other methods of separating isotopes that make use of the slight differences in the masses of the isotopes. However, the most successful methods of separating isotopes are those based on diffusion processes and the principle of the mass spectrograph.

26-8 Solar Energy

Lest we become breathless in our wanderings among the intricacies of nuclear energy and isotopes, let us turn to the most common form of energy that mankind has at its disposal—solar energy. All of us are well aware of the energy that pours out continuously from the sun, heating the earth, warming the atmosphere, and making day of the night. A weekend trip to the beach will soon remind us that the energy of the sun is prodigious, if not unmerciful, depending on how much of it we allow ourselves to capture.

The solar energy that falls on 40 square miles of the earth's surface would, if properly harnessed, supply all the needs for energy in the United States. The energy falling on each square mile is equivalent to 3600 tons of coal per day. It has been said that the solar energy falling upon the land area of the United States is some 2000 times greater than the total amount of energy we now use for fuels and food. But the problem is how to capture and store this direct energy from the sun. We now use only a small fraction of it, for we rely mostly on the sun's energy that has been stored up in fossil fuels—coal, oil, and gas. Only a small part of our energy needs are supplied by waterpower, which draws upon solar energy indirectly.

Various attempts to harness solar energy have met with only partial success. Solar engines, run by steam that is generated with solar energy, have been constructed. With large mirrors focusing the sun's rays upon a boiler, it is theoretically possible to produce an engine that turns out 50 horsepower per acre in Arizona. Houses have been heated with solar energy, but the cost of the equipment required to hold the sun's heat through the night and cloudy days rules out the economical use of solar energy for this purpose at present. Efforts to convert solar energy directly into electrical energy through semiconductors have succeeded, with enough energy being captured on a few square inches to operate a portable radio. Studies have been made of artificial photosynthesis, photochemical reactions, and synthetic fuels

with the intention of storing up solar energy for useful purposes.

But where does this vast amount of energy come from? How can the sun continuously pour out energy at the rate of three hundred thousand billion billion (3×10^{23}) horsepower? The sun has been producing energy at this rate for at least 3 billion years, because we can safely assume that during the geological past, the sun has produced an almost constant stream of energy. A rather small decrease or increase in this rate would either freeze our oceans or boil them away, something that apparently has never happened. And the sun promises to continue giving forth its energy for many more billions of years. It is thought that the sun generates its energy from a process with which we are already acquainted—nuclear fusion.

We know that the sun is composed of elements that are familiar to us. The sun's light, analyzed by the spectroscope, reveals that the sun consists primarily of hydrogen, helium, carbon, nitrogen, and oxygen, and there are not enough of the heavier elements to produce the huge energy of the sun through radioactivity and fission. Since we can examine only the light coming from the surface of the sun, we must depend on the density, the pressure, and the other characteristics of the sun to derive a picture of the source of solar energy.

It is through the study of this problem by Hans Bethe in America and Carl von Weizsäcker in Germany that a cycle of chain reactions among carbon, nitrogen, hydrogen, and helium was discovered. In this chain reaction the carbon and nitrogen are regenerated; they may be called catalysts of a sort. In effect, the process involves primarily the transformation of hydrogen into helium at very high temperatures. On the basis of the sun being composed of one per cent carbon, the rates of this transformation at 20 million degrees predicted by Bethe coincides exactly with the actual rate of production of solar energy. Each step in the complex chain of reactions has been studied on a small scale with man-made devices such as the cyclotron, so we may say with reasonable confidence, that the net reaction in the sun is:

$$4_1H^1 \rightarrow {}_2He^4 + 2 \text{ positrons} + 27 \text{ mev}$$

The two positrons are immediately annihilated by reacting with the two electrons left over from the four hydrogen atoms, since the helium atom requires only two electrons:

$$2 \text{ positrons} + 2 \text{ electrons} \rightarrow 2 \text{ gamma rays}$$

In this last reaction the gamma rays, each at 0.51 mev, are emitted in opposite directions. This peculiar reaction can be conducted in the laboratory. The small charged particles simply disappear, their masses being converted into an equivalent amount of energy.

The reactions that occur in the sun should not be compared to those that occur at ordinary temperatures and

pressures. The interior of the sun is somewhat like a gas, but at extremely high pressure. The particles are in rapid motion, but unlike an ordinary gas that is composed of molecules, the sun's gas consists of atoms whose electronic shells have been stripped off. Matter in the sun, then, consists of bare atomic nuclei mixed with unattached electrons. These naked nuclei are unprotected by their usual electronic shells, hence they may collide directly with each other. It is these many direct collisions between atomic nuclei that make it possible for the energy-producing chain of reactions to proceed. This is a much more efficient process than the scientist achieves in the laboratory when he bombards atoms with alpha or beta particles. In this case the atoms are protected by their electronic shells, and only rarely does a bombarding particle successfully penetrate to the nucleus and transform it into another element.

26-9 Summary

Nuclear energy, in contrast to other forms of energy, is unique in its magnitude, and its vast power is attributed to the exchange forces or binding energies among the particles that compose the atom's nucleus. The changes that occur in the nucleus are extremely rapid and involve very small distances.

Neutrons are important in many nuclear energy processes. Under suitable conditions, they are readily picked up by a target atom, and they are produced in several transmutation reactions as well as fission reactions.

Nuclear fission is the creation of an unstable, heavy nucleus that splits into two nuclei approximately half the size of the original nucleus. Fragments, mostly neutrons, and great quantities of energy are simultaneously released. The fission process gives experimental proof of the mass-energy equivalence law, $E = mc^2$.

The nuclear reactor provides a means of controlling the rate of fission, and thereby makes peacetime use of nuclear energy possible. The reactor also supplies the particles and energy required to carry out transmutation reactions, and for this reason it has great research potential for the physical scientist.

The fusion of nuclei is the process of building up larger nuclei from smaller ones. The most common nuclear fusion reactions involve the isotopes of hydrogen, although it is conceivable that fusion could be accomplished with several of the lightest elements.

The combination of fission of U-235, fusion of hydrogen, and fission of U-238 constitutes a three-stage military weapon that has probably been tested.

Among the various methods for separating isotopes, which is a critical step in most nuclear energy processes, are gaseous diffusion and mass spectroscopy.

Solar energy is essentially nuclear energy produced when hydrogen atoms are converted into helium.

EXERCISES

A. KEY TERMS TO REMEMBER

chain reaction	nuclear energy	solar energy
deuterium	nuclear fission	thermal neutrons
exchange force	nuclear fusion	thermonuclear process
moderator	reactor	tritium

B. QUESTIONS ABOUT NUCLEAR ENERGY

1. What is the purpose of the graphite and the cadmium in a nuclear reactor?

2. Why does carbon slow down high-energy neutrons more effectively than does iron?

3. Describe the chain reaction that occurs when a neutron is captured by the nucleus of a uranium-235 atom.

4. Compare the uranium-235 fission reaction to the hydrogen fusion reaction. How is each triggered, can they be controlled, what relative energies are released?

5. Why is there only a small possibility of setting up a nuclear chain reaction among either the lightest or heaviest elements as they occur in nature?

6. List the various methods by which uranium-235 may be separated from uranium-238.

7. If nuclear energy were to become readily available for industrial and peacetime uses (in the form of cheap electrical energy), what do you think would be the effects upon your life?

8. Which method of separating isotopes would you use if you were interested in a large volume of products? in a highly pure sample of an isotope?

9. What is the primary source of the energy that the sun produces?

10. Describe the conditions that probably exist within the interior of the sun.

11. In terms of nuclear stability, determine which element in the following pairs is more stable:

$_{3}Li^{7}$, and $_{4}Be^{9}$
$_{1}H^{1}$ and $_{1}H^{2}$
$_{2}He^{4}$ and $_{36}Kr^{84}$
$_{6}C^{12}$ and $_{6}C^{13}$

12. Explain why it is a simple matter for an atomic nucleus to absorb a neutron, but why it requires elaborate apparatus to cause a nucleus to absorb a proton.

13. Suggest an experimental procedure by which you could distinguish between alpha and beta particles.

14. Using radioactive isotopes derived from a nuclear reactor, tell how you might study (a) the wear of a metal bearing in a machine, (b) the uptake of sulfur by a plant, and (c) the metabolism of iodine in the human diet.

15. What is the essential difference between a neutron at 1 mev and a thermal neutron? Why does a 1 mev neutron refuse to be absorbed by the atomic nucleus?

16. Why need there be little concern about natural uranium undergoing a chain fission reaction on the earth's surface?

C. PROBLEMS ABOUT NUCLEAR ENERGY

1. If 0.1 per cent of the mass of the uranium in a fission bomb is converted into energy, how many joules of energy would be released by a 10-kg uranium bomb? Ans. 9×10^{14} joules.

2. If 1 kg of TNT gives off 6×10^6 joules upon detonation, what do the products weigh? Assume that the energy is derived from the loss of mass in this reaction.

3. One million electron volts equal 1.6×10^{-13} joules. Calculate the total energy released by 200 grams of deuterium and 300 grams of tritium when they undergo fusion. Hint: The equation for this reaction represents the energy evolved by single atoms of hydrogen.

4. On the basis of mass changed involved, how many joules of energy are released by the conversion of approximately 4 grams of hydrogen to helium in the sun?

5. A megaton bomb is equivalent to the explosive force of 10^6 tons of TNT. Estimate the amount of hydrogen (as H^2 and H^3) needed to make such a bomb.

Twenty-Seven

The Universe and Cosmogony

On a clear, moonless night, away from the lights and smoke of the city, we may gaze upward into the sky and experience the grandeur of the heavens. We may become engrossed in the myriads of tiny specks of light, some bright and many dim, that dot the inky depths of an infinite background. Some stars form patterns with which we may already be familiar. Other stars may appear as strangers. There are white stars, red stars, and blue stars. As we view the majesty of the night skies our curiosity grows—no wonder that man first turned to the sky in his study of the physical world about him. As an introduction to the modern concepts of the structure of the universe, let us become acquainted with the star groups called *constellations* that give us a kind of roadmap for the night sky.

We see several groups of stars in the night sky that may remind us of certain shapes or forms. Mankind for many centuries has described these shapes in terms of animals, birds, and a variety of mythical heroes. Some of these figures may appear a bit fantastic, yet with a little imagination we may visualize many of them in the constellations.

27-1 Constellations

In the northern sky, situated in an array around the North Star, are several groups of stars that we may observe almost the year round (Fig. 27-1). These *circumpolar* groups make an annual trip around Polaris, the North Star, as the earth makes its yearly trip around the sun. If we were at the North Pole, we would look straight upward at Polaris. At other

Fig. 27-1. The circumpolar stars.

locations in the Northern Hemisphere, Polaris is elevated from the horizon at an angle identical to the latitude of the observer. Thus, at 45 degrees latitude (halfway from the equator to the pole), Polaris is elevated 45 degrees, or halfway between the horizon and the zenith.

All of us will easily recognize the Great Bear (Big Dipper) in the northern sky. It is a bit more difficult to find the Little Dipper, the Dragon, and Cassiopeia. But with a little practice, using Fig. 27-1 at night out of doors, we may soon become acquainted with these groups of stars. Some of the most interesting groups, however, are in constellations that are viewed only during certain seasons of the year. For example, Orion the Hunter is best observed during the evening hours of the winter months when it can be seen moving from night to night in a westerly direction. After becoming ac-

Fig. 27-2. A portion of the sky in winter, showing Orion and other major constellations. (To use this map properly, hold it over your head.)

quainted with this constellation (Fig. 27-2), we should have no difficulty in finding it again—there is no other arrangement of stars like it in the sky. The stars themselves in Orion are of special interest, for if we look closely we may see that they are not all the same color. In his sword, the middle star is not really a star but a mass of fuzzy light that shows up when examined with a small telescope or even a pair of binoculars. Near Orion is the constellation Canis Major (Big Dog), too low in the southern sky for us to see well, but containing the star Sirius that outshines all the others in the sky. On the other side of Orion is Taurus the Bull, which contains a fine collection of six faint stars known as the Pleiades.

The names of at least sixty constellations have come down to us through the years. Besides enhancing our pleasure in enjoying the stars, these groups of stars have served the astronomer as a set of reference points. He is like the earth-bound tourist who studies a road map to find the location of a town, say, Springfield, in the state of Illinois. On a "sky map," the astronomer may speak of a star, say, Arcturus, as being in the constellation Boötes (the Herdsman). As we become more acquainted with the star groups and the seasons of the year in which they appear, we will enter upon a new and exciting experience that has been shared among men for centuries. The behavior of stars is more reliable than almost anything else in all nature, and an acquaintance with them is like standing on the threshold of the universe itself.

On the other hand, the constellations are not true groups of stars, for some of the stars in a particular constellation may be much closer to us than others. It is only an optical illusion that they appear to be in the same plane. Even in the great Milky Way, which is a band of billions of stars, the stars are scattered near and far from us. And, if we could compare the position of the stars in a constellation, we would find that they move, sometimes as a group, but often in different directions. The stars in the Big Dipper, for example, are moving away from one another, so that eventually this group of stars will have no resemblance to a dipper.

27-2 The Nearer Stars

Next to the sun, which is truly a star and very similar to many stars we see in the sky, the brightest star is Sirius, in the Big Dog. This star is 26 times brighter than our sun, but being 8½ light-years away it appears only as a bright pinpoint of light. A *light-year* is the distance traveled by light in one year—approximately 6 million million miles. Because of the great distances involved in the study of the stars, it is a matter of convenience to express these distances in light-years. The nearest star, Centauri, is only 4 light-years from us. The brightness of a star is measured in terms of

magnitude, and the difference between one magnitude and the next is a factor of 2½ times in brightness. We can see up to a sixth- or seventh-magnitude star with the unaided eye. The brightest stars are of a magnitude of 1 to 0, although Sirius is so bright we have to use a negative figure, —1.58, to designate it. In the Big Dipper five of the seven stars are of second magnitude, so it is no wonder that this constellation is easy to find. Table 27-1 lists several of the brightest stars and a few of their descriptive properties.

TABLE 27-1 THE BRIGHTEST STARS

Star	*Constellation*	*Magnitude*	*Color*	*Distance in light-years*	*Size* [1]	*Velocity* [2] *in mi/sec*
Sirius	Big Dog (Canis Major)	—1.58	blue	8	2	4.6
Centauri	—	0.06	yellow	4	—	13.8
Vega	Lyrae	0.14	blue	23	2.5	8.6
Arcturus	Boötes	0.24	orange	32	25	3.2
Rigel [3]	Orion	0.34	blue	545	35	—14.7
Betelgeuse	Orion	0.92	red	300	248	—13.0
Aldebaran	Taurus	1.06	orange	54	34	—33.6

[1] Size (radius) is given in comparison to the sun.

[2] Velocity is given in miles per second approaching the sun, or receding from the sun (—).

[3] The star, Rigel, bears the distinction of being the brightest star known to us. However, because of its great distance it appears no brighter than several others in the night sky. It has been estimated that Rigel puts forth 14,000 times as much light as our sun.

The distance to the nearer stars is measured by the same method that a surveyor uses—triangulation. If a surveyor wishes to measure the width of a river, he carefully measures the length of a base line on one side of the river and then finds the angles between the ends of this base line and some point on the opposite shore. In astronomical surveying, we need as long a base line as possible in order to obtain sufficient accuracy in our measurements. If a star is observed from the opposite sides of the earth, the diameter of the earth would be the base line. The longest and most dependable base line, however, is obtained by making observations of a star at six-month intervals. The base line now becomes the diameter of the earth's orbit around the sun, a distance of twice 93 million miles (Fig. 27-3). With this method, the distances to as many as 5000 of the nearer stars have been measured.

Once the distances to the nearer stars are known, we may derive another important character of the stars—*absolute magnitude.* Our sun is of —26 apparent magnitude, its nearness being responsible for its brilliance. Many stars are far brighter than our sun, but because of their greater distance

from us they appear as dim stars. If we know the distance to a star, as well as its relative brightness compared to our sun, we can calculate the actual or absolute magnitude of a star compared to our sun.

We saw in Chapter 20 that the color of a star is an indication of its surface temperature. We know that a piece of iron is much hotter when it is white-hot than when it is red-hot. Similarly, the hottest stars emit light that is strongest in the violet end of the spectrum. In general, we may say that the temperature of a star varies inversely with the wave length of the light in emits. Thus, the coolest stars are red, the hottest stars are blue (Fig. 12-6). Although there are several complexities in the relation between the color and the temperature of the stars, it is through color studies that the temperatures are best measured. From the Planck relationship (Chapter 12) we saw that the amount of energy at different wave lengths is a function of temperature. Hence, the most accurate measure of a star's temperature will involve a study of all the light that the star emits. When we say that a star is red, we merely mean that the red wave lengths of light predominate, and a spectroscopic examination of the rest of the colors must be made before we can state definitely the star's temperature. The hottest stars have surface temperatures of almost half a million degrees, while the cooler stars may have temperatures of only 1700°C. The internal temperature of the stars is quite another matter. Because there is no direct means of observing the interior of the stars, we can only conjecture about their temperature. If a star's source of energy is similar to the sun's, its interior temperature is probably about 20 million degrees.

There is an enormous variety among the stars. The diameter of the largest is 300,000 times as large as the diameter of the smallest. There is a thousand times difference in their masses, and their densities vary from that of almost a vacuum up to tons per cubic inch. The companion star of Sirius is not much larger than the earth, yet is 200,000 times heavier in mass. All the stars are completely gaseous, and they are held together by gravitational forces even though they are sometimes greatly distended by the energy of radiation that pours out from their interiors.

But most of the stars fall into a nice sequence when arranged according to temperature, brightness, and size. This so-called main sequence of stars (Fig. 27-4) contains many of the common stars we see at night as well as our sun. Outside this group of main-sequence stars are several super, or giant, stars, which are large, low-temperature, low-density stars that are far less common than the main-sequence stars. To the lower left of Fig. 27-4 is a remarkable group of stars, the dwarfs. These are high-temperature stars with incredibly high density and small size; some of them are billions of times more dense than the giant stars.

The stars appear fixed in the sky because from night to night and from year to year they seem to occupy exactly the

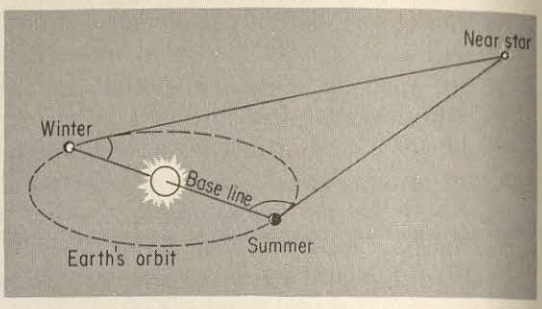
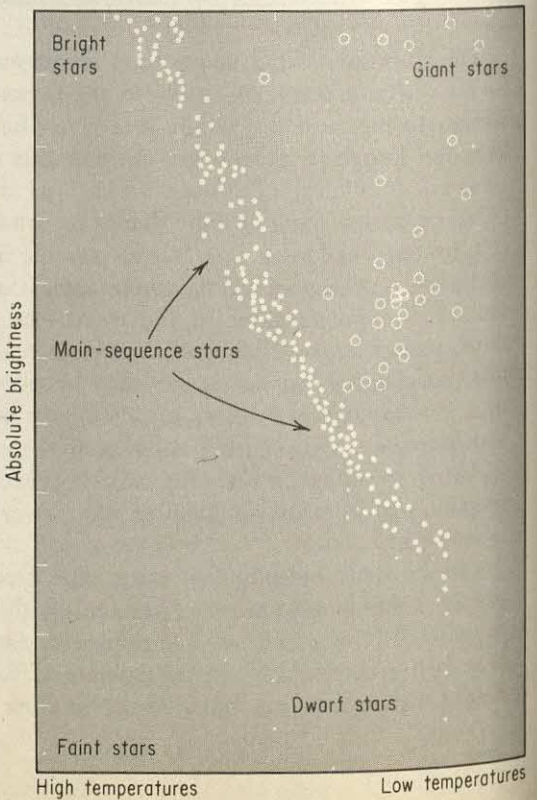


Fig. 27-3. Using the earth's orbit as a base line in the triangulation of a star.

Fig. 27-4. The so-called Russell diagram of the principal types of stars: magnitude vs temperature.



same positions. However, when we compare photographs of certain star groups made several years apart (Fig. 27-5), we can be assured that the stars do move. Because of their remoteness, the movement of the stars appears small, but repeated observations leave no doubt that there is a high degree of unrest among the stars.

In addition to the movement of stars across our field of vision, they may be moving either toward or away from us. This kind of motion is called *radial motion*, so named because the movement is along the radius of a circle relative to an observer at the center of the circle. The radial motions of several stars are listed, under velocity, in Table 27-1. To measure the radial motion of a star, we use the Doppler principle, mentioned in Chapter 10. You will recall that the wave length of a wave train may be increased or decreased if the source and observer are moving toward or away from each other. Thus, if a star is moving toward us, its light will be shifted toward the violet or shorter wave lengths of the spectrum. If a particular spectra line for, say, excited hydrogen atoms ordinarily falls at 6563 A, and the same line in the light from a star falls at 6540 A, then the star is moving toward us. From the velocity of light and the Doppler principle we may calculate the velocity at which the star is moving.

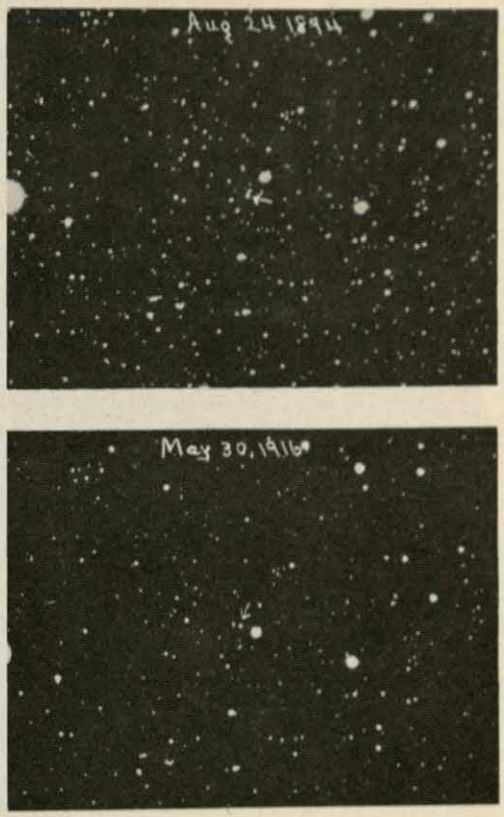


Fig. 27-5. Evidence that stars move is shown in this pair of photos made 22 years apart. The change in position of Barnard's star in the constellation Ophiuchus is marked. (Yerkes Observatory photograph)

27-3 Variable Stars

The surprise packages of the night skies are the *variable stars*, which are so termed because they change in brightness from hour to hour, or month to month. Polaris undergoes a 10 per cent change in brightness each four days. Some giant stars vary in brilliance by as much as 10,000 times in their cycle. Variable stars are found among the nearer stars as well as among the farthest ones we can see with the biggest telescopes. Many hundreds of them have been observed.

There are several reasons for the variations in a star's brightness. The simplest kind of variable star is called an *eclipsing variable*, which is merely a pair or cluster of stars that circle around each other, much as planets travel around the sun. If the plane of revolution is viewed from the edge, one of these circling stars cuts off the light from another star in a simple eclipse that is particularly noticeable when the stars involved are of different brightness.

But there is another group of stars, far more interesting than the eclipsing variables, that change in intrinsic brightness. These are called *Cepheid variables*, and for some mysterious reason their light undergoes a pulsating brilliance over periods that vary from less than half a day to several months. Enough Cepheid variables have been observed among the nearer stars to provide us with a very valuable measuring stick. It has been found that the period of pulsa-

tion of a Cepheid variable is related to its absolute magnitude. You will recall that the absolute magnitude of the nearer stars can be found if we know the distance to them. Every Cepheid variable with the same period seems to have the same absolute magnitude, and this is all the astronomer needs. As he peers into the limitless depths of the universe and observes a Cepheid variable, he determines its period of pulsation, and this period tells him the absolute magnitude of the star. A comparison of the absolute magnitude with the apparent magnitude readily gives the distance through which the light from this star traveled. Distances of upward of a million light-years have thus been measured.

Another group of variable stars has a pulsation period that is much more extended. These long-period variables change in brightness by a factor of hundreds of times, and their periods average 280 days. These stars, sometimes called *Mira stars,* are among the largest known. The biggest one has a volume over 25 million times that of our sun, though it is a much cooler star than the sun. This particular giant has a density of only one-thousandth of that of the air we breathe, which is almost a vacuum in the ordinary sense. How so tenuous a body at a relatively low temperature can emit light is indeed a mystery of the universe.

As we saw in Chapter 23, the energy being radiated from the sun seems to have remained remarkably constant. There are, however, variations in the brightness of our sun. Some of these variations are local in nature and are particularly noticeable on certain areas of the sun's surface. Both sunspots, which last on an average of a few days and occur in regions that have previously been bright patches, and solar prominences are indications of the variable nature of the sun. Even though the sun shows no noticeable variation as a whole, we have seen that it is possible that long-period variations may have been responsible for the ice ages and the glaciation in the geological past.

At any rate, the structure of the sun's surface and the observable variations in many stars reveal an element of change and unrest in the universe. What we may once have thought to be static and unchanging is passing through a continuing process of alteration and change. This may cause us to wonder about the origin of the universe and the process of change through which it is now going, and this matter we shall consider in detail shortly.

27-4 Novae and Supernovae

A *nova,* meaning a "new star," is an ordinary-appearing star that suddenly flares up to an amazing brilliance. What may appear as a common star of no greater absolute magnitude than our sun suddenly increases thousands of times in brightness. Then, after a while the nova fades to its original status as an ordinary star. Several novae have been studied

carefully. Let us consider one of these as a typical example.

The Nova Aquilae, prior to its outburst in 1918, was an ordinary variable star of bluish white color. Then, in the course of only a few hours, it burst forth, sending out large jets of glowing material. Spectra observations showed that the surface of the nova exploded outward at the rate of 1050 mi/sec. Like a huge Roman candle, two flaming sprays came outward, one almost directly toward us and the other in the opposite direction, and several smaller jets produced a "smoke ring" of luminous gas, all within about six hours. This mass of ejected matter continued to move outward, and even today it continues its expanding pattern, apparently no longer attracted by gravitation to the parent star, which has long since died down to its original existence as an ordinary variable star. In Fig. 27-6 the expanding, glowing gases of a nova are photographed fifty years after the original explosion occurred.

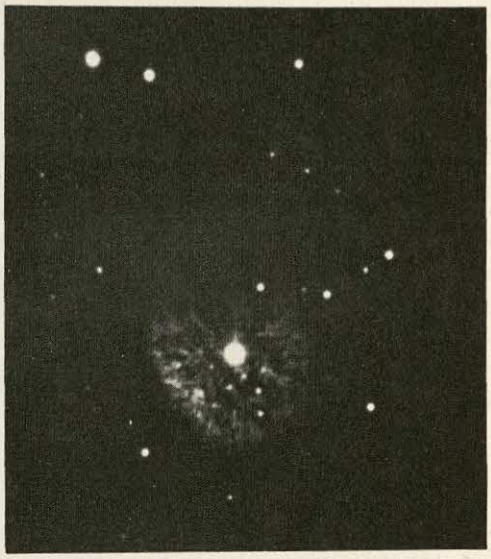


Fig. 27-6. Nova Persei, 1901, photographed 50 years after its original eruption. (Courtesy Mount Wilson and Palomar Observatories)

Occasionally a star that appears as an ordinarily nova bursts forth into a gigantic explosion that literally destroys the whole star. These stellar catastrophies are called *supernovae*. The explosion of an ordinary nova may reduce the nova's mass by only a fraction of one per cent, its mass loss being equivalent to the energy it releases. But with supernovae the process is much more energetic. What little we know about supernovae indicates that they are stars that have used up most of their hydrogen, so that they no longer have enough fuel to operate by a usual nuclear energy reaction. These fantastic outbursts, in which a major portion of their mass is thrown off, may then be considered as a closing out of the star because of hydrogen bankruptcy. Supernovae, however, are too rare to permit a regular study of their habits. The most recent one that occurred close enough for observation was in 1885, prior to the time that we had sufficiently accurate spectroscopes to study their spectra. These magnificent spectacles of the universe, therefore, remain a puzzle.

Fig. 27-7. The Crab Nebula in Taurus is the remains of a supernova explosion of A.D. 1054. (Courtesy Mount Wilson and Palomar Observatories)

The Crab Nebula (Fig. 27-7) is a good example of a supernova. A *nebula* is a mass of cloud—a thin, almost transparent yet glowing region—which at a distance may appear as a single star. Upon magnification it may show structure, although other stars may be seen through it. From a series of photographs and spectra studies, the Crab Nebula was found to be expanding away from its center at the rate of about 70 million miles a day. Hence, from its rate of expansion we can calculate back to the time when it started. The original outburst must have happened about 900 years prior to its present status. Interestingly enough, Japanese and Chinese records show that a bright new star was observed in the constellation Taurus on July 4, 1054. It seems practically certain that the nebula we now see (Fig. 27-7) is the debris of a supernova that burst in a colossal explosion at that time.

In addition to the visible light radiated from the Crab

Nebula, in recent years other types of radiation have been detected. Long-wave radiation in the radio frequency range is emitted from this and many other parts of the universe. These radio signals (Fig. 29-3) provide a clue to the composition of the star, and oftentimes these signals come from otherwise dark regions of the universe. At a wave length of about 21 cm, these signals reveal the presence of cool, neutral hydrogen atoms. These so-called *radio stars* provide information about the stars and interstellar space in addition to that obtained from the visual light that comes to us from the stars. The Crab Nebula also emits X rays and gamma rays, and these types of radiation provide clues to the composition and temperature of this part of the universe.

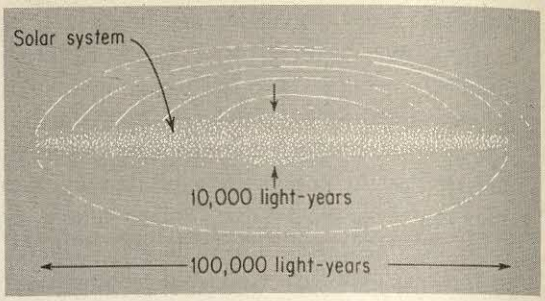


Fig. 27-8. A cross-sectional diagram of our galaxy.

27-5 Our Galaxy

A *galaxy* is a complex system of stars, star clusters, and gas and dust that is irregularly distributed within a rather well-defined volume. Our galaxy is difficult to visualize because we are situated within it, and as we gaze at the Milky Way we are looking through the thick portion of it. It is shaped somewhat like a sofa pillow, 100,000 light-years in diameter, 10,000 light-years thick at the center (Fig. 27-8), and contains most of the stars we see in the night sky.

The solar system is situated in our galaxy about two-thirds of the way from the center to the edge. Most of the stars and constellations that we see at night with our naked eyes are immediately adjacent to our galaxy. Since in the winter we are looking outward from the galaxy and in the summer we are looking toward the center of it, the Milky Way appears much brighter in summer than in winter. When we look either to the left or to the right of the Milky Way, we are looking out into the voids beyond the edges of our galaxy. These edges are not abrupt; rather they taper off with fewer and fewer stars until there is nothing.

Fig. 27-9. A spiral nebula, Messier 51, showing the spiral structure of a galaxy when viewed face-on. (Courtesy Mount Wilson and Palomar Observatories)

Our galaxy is not uniformly populated with stars. In fact, if we were to move outside our galaxy we could see that it is composed of huge spiral arms that radiate out from the center. The structure of these arms may best be visualized by actually observing another galaxy outside our own (Fig. 27-9). Through a careful study of the number of stars in various regions of the Milky Way and of radio radiation from the atomic hydrogen concentrated in the spiral arms of the galaxy, we can show approximately the spiral structure of our galaxy (Fig. 27-8).

In addition to stars and star clusters, our galaxy contains a tremendous amount of gas and dust. As we look at the Milky Way, we may see several irregular bands where far fewer stars appear. These dark patches of the sky are areas that contain enough dust and gas to blot out the stars behind, and where the stars in the Milky Way are thickest, these obscuring clouds are the most common.

Since quite often the mass of dust and gas in the space between the stars is as much as the mass of the stars themselves, the astronomer is much concerned with the composition and behavior of all this dust and gas. For one thing, the light that comes to us from distant stars must pass through billions of miles in which there is considerable material. Even if there is only one molecule of gas per cubic centimeter, the light will be sufficiently distorted to require compensation. Just as our sun's light appears more red when it passes through a dust cloud, so starlight appears more red when it passes through interstellar dust and gas. The absorption of light by this dust indicates that a major share of it is in the colloidal range of sizes. However, there remains much work before a complete understanding of all this dust and gas is achieved.

27-6 Galaxies and the Universe

Our own galaxy, huge as it is, is not the end of the universe. Beyond our galaxy and isolated from one another in space are perhaps a billion more galaxies that can be seen with our present telescopes. There is no way of knowing how many galaxies exist beyond the range of the largest telescope because the universe apparently extends without limit. Some of these galaxies are spiral-shaped like our galaxy (Fig. 27-10), whereas others are irregular. The galaxies nearest us appear as faintly visible scraps of the southern Milky Way. They were first recorded by Magellan in 1519 while on his trip around the globe and are thus called the Magellanic Clouds.

Fig. 27-10. A spiral nebula viewed edge-on. (Courtesy Mount Wilson and Palomar Observatories)

The distant galaxies contain thousands of Cepheid variables, which appear to behave very much like the Cepheids among our nearer stars. By knowing their periods and apparent magnitude it is possible to calculate the distance to them, and this information gives us the best method of measuring the distances to the nearer galaxies. For example, evidence from the Cepheid variable stars in the great spiral galaxy in Andromeda indicates that these stars are 1,700,000 light-years away. And this galaxy is so near that we can see it with the naked eye as a hazy speck of light in the constellation Andromeda!

The average distance between galaxies is about 2 million light-years. With a 100-inch telescope that sees out to about 500 million light-years, it should be possible to examine 100 million galaxies. The 200-inch telescope at Mt. Palomar, which reaches out twice as far, covers a volume eight times greater. With this larger instrument, up to almost a billion galaxies can be studied. Most of these galaxies are so far from us that the most powerful telescope cannot resolve them into individual stars, but their distances can be calculated from their apparent brightness. This is truly a bewildering universe—a collection, without a known end, of huge gal-

axies, each one of which contains at least a billion stars, and of stars that are scattered many light years from each other. How humble we feel as we become aware of the immensity of it all!

But the universe is not at rest, no more than are the stars that make it up. You will recall that some of the nearer stars are moving toward us, some are moving away. Within our own galaxy there is undoubtedly a great deal of turbulent motion, for its spiral structure is ample indication that there are irregular movements in it. In addition to the movements of stars within the galaxies, there is a more uniform movement on a grander scale—the galaxies themselves are gradually drifting farther and farther apart.

Here is one of the most amazing of phenomena in the realm of physical science. A study of the light from distant galaxies shows that the spectral lines are all shifted toward longer wave lengths. On the basis of the Doppler effect, this so-called red-shift indicates that these galaxies are receding from us and that the universe is an expanding system of galaxies. Although there are other possible explanations of the red-shift, it is most commonly explained as the result of an expanding universe.

The red-shift is even more evident for the farthest galaxies from us, and this fact implies that these galaxies are receding at a more rapid rate than are the nearer ones. Indeed, they may be traveling away from us at the rate of 61,000 km/sec, which is one-fifth of the velocity of light! We should

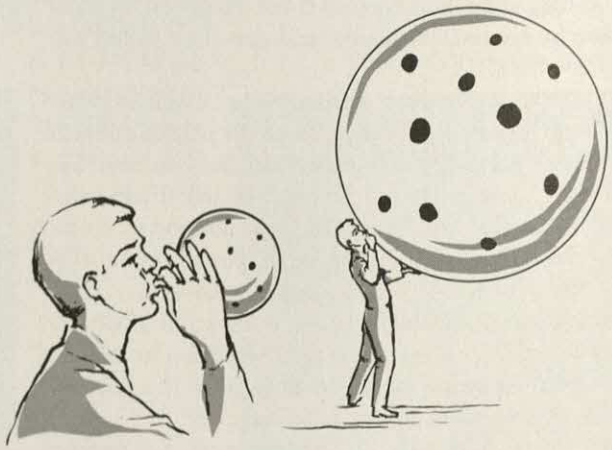


Fig. 27-11. A model of the concept of an expanding universe.

not think that all the galaxies are rushing away from us, as though we particularly repelled them, for they are all moving away from each other. We may easily grasp this idea if we inflate a rubber balloon speckled with painted polka dots, each representing a galaxy (Fig. 27-11). When the balloon is stretched, the dots all move away from each other, and this is apparently what is happening to the spacing of the galaxies.

27-7 Evolution of the Universe

We are now in a position to assess the facts that have been introduced up to this point and to make a tentative analysis of their meaning. When we are aware of the size, time, and masses involved in the universe, we are naturally driven to raise the question—when did it begin?

Scientific men have long ago given up the idea that the date of the Creation can be determined from historical records, such as the sum of the ages of the patriarchs as calculated by Archbishop Ussher. It has turned out to be much more dependable to keep time by the atoms, the sun, and the stars than to rely on the human ability to measure time intervals. Similarly, the records of time in the distant past appear to be more faithfully preserved in the atoms, the rocks, the sun, and the stars than in any man-made records that may have been passed down through countless generations of prehistoric and historic men.

Now that we have examined some of these natural records that are available to us, let us reconsider them for our immediate interest. In the study of radioactivity, we found that only those isotopes of radioactive elements that have rather long half-lives are found in nature. Apparently those that have short half-lives have long since decayed into stable isotopes. But such radioactive isotopes as uranium-238 decay slowly enough that we may calculate back to the time when a particular uranium ore was laid down as a solid portion of the earth's crust. Thus, we can date rocks, minerals, and the earth's crust.

We made a guess as to the age of the ocean by the amount of salt it contains and the rate at which salt is added each year. We can guess at the age of the moon on the basis of its radial movement; it is receding from us at the rate of about 5 inches per year. The age of the sun can be estimated from the amount of energy it puts out, its brightness, and the quantity of nuclear fuel it burns up. The age of the galaxies may be estimated simply from their mechanical movement. From Newton's first law, a body in motion continues in uniform motion unless acted upon by an outside force—the galaxies have been moving uniformly in an expanding universe. At some point in time, they must have started from a common origin when tremendous forces sent them outward on their travels. A knowledge of the energy distribution of the light from stars in our own galaxy gives us still another means of making an estimate of the age of a portion of the universe. When we assemble all these various ages (Table 27-2), we note that there is a striking similarity between them. These values are not identical, but they agree with each other in order of magnitude to an extent that assures us that the agreement is no mere coincidence.

These figures support the idea that at some time in the past, say about three or more billion years ago, things as we now know them had their origin. At this time, the earth was

probably formed, the moon began its orbit around the earth, the stars began to shine, and the galaxies began to start their lengthy wanderings into limitless space. If we accept these time measurements, we must consider that the universe is not a permanent, unchanging system but the result of some sort of evolutionary process that began some few billions of years ago. Then our curiosity naturally raises the question of how the process was started. Was it all done in a single swoop? What preceded the 4-billion year mark? If this is an evolutionary process, are stars and planets still being born? If so, what happens when a star is created? For these and many more fascinating questions, scientists are busy finding answers. But let us remember as we examine a few of these answers that they are only tentative ideas and that with more knowledge they are likely to be changed considerably.

TABLE 27-2 THE AGES OF VARIOUS PORTIONS OF THE UNIVERSE

	Age (in billions of years)
Atoms (by radioactivity)	4 to 7
Rocks (by radioactivity)	2
Earth (by historical geology)	3.3
Ocean (by quantity of salt)	1.5
Moon (by recession rate from earth)	4
Sun (by rate of energy emission)	3
Stars (average)	3
Galaxies (by mechanical movement)	1 to 10
Milky Way	2 to 5
Universe (by red-shift in star spectra)	4

How is a star born? Well, let us take a close look at those huge dust clouds that obscure various parts of the Milky Way as well as parts of neighboring galaxies. They are composed of helium, hydrogen, carbon, oxygen, nitrogen, and so on, probably in chemical combinations that are familiar to us. The particles of dust may be considered as ordinary dust particles—they scatter light, cast shadows, and exert gravitational forces on each other. The clouds may have been pushed together by the small forces of light shining from the stars (Fig. 27-12). As small clouds formed they grew because they obscured starlight, and new dust particles were driven to them. If no star comes near enough to cause a turbulence and a cloud continues to grow, it may reach a size where it has sufficient mass and density to draw itself together by gravitational forces. As the cloud begins to contract it warms up, in keeping with the compressional nature of gases as we know them, and, supposedly, these compressional forces would be strong enough to raise the temperature to that required to start nuclear energy reactions. Thus, the dust cloud is converted into a star.

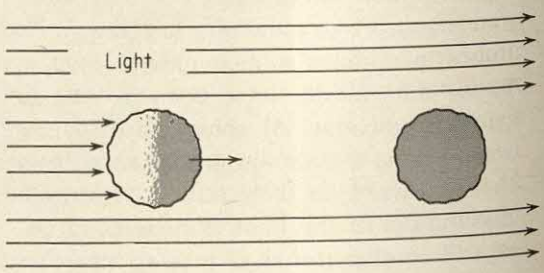


Fig. 27-12. The force of light pushing two dust particles together.

There are good indications that stars have formed well since the "beginning," because certain stars are much too bright to have existed for even as much as two billion years. That is, their brightness is so great that they would have long ago burned up their nuclear energy if they had started two billion years ago. Therefore, they must have been born later than many other stars and our sun. So convincing are the arguments of the dust cloud hypothesis that, as we shall see in a moment, they have been used also to explain the origin of the planets in our solar system.

Now let us turn to a hypothesis concerning the origin of the universe. On a grand scale, the famous astrophysicist, George Gamow, has drawn for us a most interesting picture of what happened at the beginning of the great expansion. Prior to the start of the expansion there must have been a "Big Squeeze"—a completely heterogeneous affair that consisted of a most unusual collection of primary matter (protons, neutrons, and electrons). Temperatures of possibly a billion degrees prevailed. The early stages of expansion cooled this material and encouraged the formation of atoms —particles with neutron-proton cores and electron shells. This big squeeze and sudden expansion may have occurred in a very short time, possibly an hour, and it may well have been like the center of an exploding nuclear bomb.

Then, according to Gamow, the newly formed atoms expanded for many millions of years as a uniform mass of gas pushing outward and cooling, possibly to "room temperature." Dust particles began to form, and at this stage the force of gravity began to break up the gas and dust into clouds. This consolidation of particles may seem inconceivable at first, but we must realize that the random motion of gas molecules rules out the possibility of a perfectly uniform gas structure. Occasional irregularities occur even in our own atmosphere, but their gravitational forces are too small to hold them together. Thus, the gas and dust broke up into clouds, with almost nothing between them (Fig. 27-13). These clouds may have been protogalaxies that gradually broke up into smaller clouds that may have been protostars. Within these smaller clouds, gravitational forces rapidly brought about a contraction that produced high temperatures, and the interior of these protostar dust clouds became hot enough to start nuclear reactions such as we know occur in the sun at the present time. Thus the contracting dust cloud became a luminous body—a star. Along with these protostar dust clouds there may have been much smaller clouds that condensed but never reached the high temperatures required for nuclear reactions. These smaller clouds may have produced the planets.

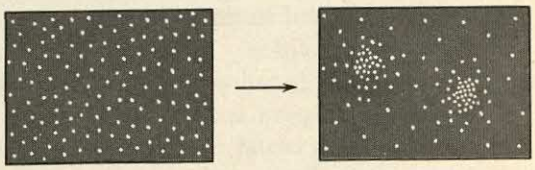


Fig. 27-13. Condensation points in an originally uniform system.

Once the stars and planets were formed, we can be fairly certain of the progress of the evolutionary process. It is the picture prior to the geologic past that is hazy and not well understood. For this reason we must view the dust cloud hypothesis and the Gamow theory as only tentative ideas.

Gamow summarizes the time factors of the various steps involved in the evolution of the universe as follows: "It took less than an hour to make the atoms, a few hundred million years to make the stars and planets, but three billion years to make man!" This conclusion reveals an important difference between physical and organic processes. Compared to the physical evolution of the universe, which required only a small amount of time, organic and biological evolution has occurred at a very slow rate.

Regardless of a person's beliefs, he must consider the vast universe with its galaxies and endless space as a miraculous creation. The unbelievable size and incredible magnitude of the universe are without an explanation. Whatever the implications for one's personal views, our present knowledge of the universe shows that we stand only upon a threshold looking out upon a great outer existence. How little is man and yet how great is the world upon which he gazes. Does not man have sufficient reason to be humble indeed?

27-8 Summary

Constellations are groups of stars that are observable with the naked eye in the night sky to which various shapes and forms, and often mythological characters, have been assigned. They provide a ready reference to certain regions of the sky as well as guideposts in earthly navigation.

Stars, including our sun, are classed according to magnitude (both apparent and absolute), temperature, and size. In such a classification, most stars fall into a main sequence, with only a few extremely large ones and very few small ones outside this sequence.

Cepheid variable stars provide a key to the measurement of distances to farther stars because their period of variation is proportional to their absolute magnitude.

A star that suddenly flares up, then fades back to its original intensity is called a nova; a supernova flares up more intensely than a nova and literally destroys itself.

A galaxy is an island cluster of stars, gas, and dust, separated by great distances from each other. The solar system is a tiny part of our galaxy, which is apparently shaped like many galaxies that can be observed at a distance. The structure of galaxies, with their spiral arms, indicates a great degree of movement.

A study of galactic structures—their movements and distances apart—gives considerable basis for theorizing about the origin of the universe. The red-shift of star spectra indicates that the universe is probably expanding, and, along with the ages of various portions of the universe, we have considerable evidence to support our ideas about the process by which the universe has reached its present status.

EXERCISES

A. KEY TERMS TO REMEMBER

Cepheid variable star	light-year	nova
constellation	main-sequence star	radio star
Crab Nebula	Milky Way	red-shift of star spectra
eclipsing variable star	Mira star	supernova
galaxy	nebula	

B. QUESTIONS ABOUT THE UNIVERSE AND COSMOGONY

1. Using the diagrams of the constellations in this chapter, locate as many of them as you can in the sky on a clear, moonless night.

2. If you are an amateur photographer, take a picture of the North Star with a time exposure of at least half an hour. Explain the apparent motions of the stars that are recorded in this manner.

3. List the different kinds of variable stars.

4. What is the relationship between the color of a star and its temperature?

5. Explain the method of measuring astronomical distances by observing Cepheid variable stars. What assumptions are the astronomers making?

6. Is it possible to have a longer base line than the earth's orbit for measuring stellar distances by the triangulation method?

7. Explain why the Milky Way appears much brighter in summer than in winter.

8. Bearing in mind the structure of a typical galaxy, what do you suppose would happen if two galaxies collided with one another?

9. Discuss one of the modern theories concerning the process by which the universe was formed.

10. In the study of the movement of stars in a galaxy, the spectra of the stars in the galaxy are examined. Which view of a galaxy would be better to study for this purpose, an edge-on view or a face-on view? (Compare Fig. 27-9 and Fig. 27-10).

11. Consider the possibility that the Star of Bethlehem might have been the result of a natural phenomenon described in this chapter. In view of the Biblical account that this star shone brightly for several weeks and then passed away, what might it have been?

12. From Table 27-2, estimate the age or time at which Gamow's "Big Squeeze" and sudden expansion might have occurred.

13. Is it possible that a universe exists outside our known universe? Keeping in mind the velocity of light, how would it be possible to know about such other universes?

14. Explain the differences in the ages shown in Table 27-2. For instance, why is the calculated age of the oceans considerably

smaller than the calculated age of the moon even though they may be actually the same age?

15. We probably know more about the structural shape of other galaxies than our own. Why?

C. PROBLEMS ABOUT THE UNIVERSE AND COSMOGONY

1. What would be the wave length of the spectra line of hydrogen observed in the star Aldebaran that ordinarily is 6563 A? See Table 27-1 for the radial movement of this star.

2. Recalling that light intensity falls off with the square of the distance, approximately how many times brighter is Betelgeuse than Aldebaran in absolute magnitude?

3. If our days suddenly became one-thousandth of a second longer, how much later (or earlier) would a predicted event, such as an eclipse of the sun, occur one thousand years in the future?

4. If a nova changes in apparent brightness from magnitude 6 to magnitude 1, how many times brighter has it become?

Twenty-Eight

The Availability of Energy

We encountered the concept of energy early in our study of physical science. In Chapter 2, energy was defined as the capacity for doing work, and work was defined as a certain force operating through a certain distance. These definitions give us a mechanical concept of energy, but we have, of course, encountered other forms of energy—heat, chemical, nuclear, and radiant energy.

Although the material composition of a system is often of primary interest to us, there are other properties that may be equally important. The temperature, pressure, heat content, motion, and electrical charge may be of more immediate concern than the chemical composition of a substance. For example, we are usually more interested in the energy of a rifle bullet traveling at 3000 ft/sec than in its composition. When a thousand tons of rock slide downhill in a mighty avalanche, it is the conversion of potential energy to kinetic energy that particularly concerns us. We burn gasoline in our automobiles not for the carbon dioxide and water vapor that issue from the exhaust pipe, but for the energy that is released by the combustion process. When we burn fuels to heat our homes, we purposely discard the combustion products but conserve the heat energy that is evolved.

But how do we obtain energy when we want it? Are we merely content to drive into the nearest gasoline station and say "fill 'er up"? Why do we use gasoline for fuel in our auto? Why not use water or some other material that would be much cheaper? The answer to this is obvious to most of us—gasoline is a readily available source of energy whereas water, unless it is permitted to fall through a hydroelectric

generator, is almost useless as a direct source of energy. We use water in a steam boiler to drive the piston of a steam engine, but we must first supply the energy of burning coal to the water to convert it into steam. The water serves merely to transport the heat energy.

Now we turn our attention to some of the basic aspects of energy in general—where do we find it, how can we control and use it, and what factors determine when it is available to us?

28-1 Sources of Energy

Perhaps the most "primitive" source of energy that is still common to all of us is biochemical energy—the energy produced by the muscles of a man or an animal. As we consume foods, their stored-up chemical energy is converted into heat and muscle energy. Through many ages man has capitalized upon his own or animal biochemical energy to do his work, and this type of energy continues to be important even in our modern day, for who among us would wish to be without our own muscle energy to propel us about our daily activities?

In modern times, less and less animal power is used to supply our energy needs, while more and more mechanical power is put to use. In this country we employ 50 times as much energy to run our machines as we do to run our bodies. In order to supply the tremendous amounts of mechanical power required by our modern civilization, we draw upon the natural storehouse of energy in the earth's crust. The most common sources of this energy are the fossil fuels: coal, oil, and natural gas. These fuels were created by complex geological processes that are not fully understood. Apparently, dense, lush forests of abundant plant life were deposited and converted into an organic sedimentary rock that became the deposits of coal in the earth's crust. Oil and natural gas fuels, as noted in Chapter 16, apparently resulted from marine micro-organisms that operated upon organic sediments and were later capped by an impermeable layer of younger sedimentary rock. The whole process of petroleum formation remains much of a mystery, although it is generally agreed that its primary origin was plant and animal life in the ocean.

There is considerable variation in the estimates of the quantities of the fossil fuels ultimately available in the United States (Table 28-1). The world fuel reserves are estimated to be approximately three times those shown for the United States. Note that coal is far more abundant than oil and gas, and we must expect that as the reserves of oil and gas are used up, we will need to fall back on coal, unless other sources of energy are developed. This will require considerable changes in either the manner in which we use the fuel or in its preparation. Because of the popular trend to-

TABLE 28-1 THE WORLD'S FOSSIL FUEL SUPPLY

Fossil fuels	*Occurrence at accessible depths (metric tons)*	*Exploitable by present technology (metric tons)*	*1962 Rate of world production (metric tons)*	*Probable years of supply in exploitable reserve*
Coal	1.5×10^{13}	4.7×10^{12}	2.7×10^{9}	1755
Petroleum	7.0×10^{11}	5.0×10^{10}	1.2×10^{9}	41
Natural gas	1.9×10^{11}	1.8×10^{10}	5.0×10^{8}	36

Compiled from Elmer H. Pehrson, "Minerals in National and International Affairs," *Economics of the Mineral Industries*, 2nd ed. (New York: American Institute of Mining, Metallurgical and Petroleum Engineers, 1964), p. 526.

ward liquid and gaseous fuels, we will need to find cheap methods of converting coal into one of these kinds of fuel, which will pose many problems for the fuel scientist and engineer.

Most of these scientists and engineers have been quite unwilling to face the possibility of low future petroleum reserves. Although the annual production of petroleum in the United States has steadily increased up till now, it is expected that in not too many years our production will gradually fall off as new oil reserves become scarcer. The search for new deposits is intense—yet only one in nine of the wells drilled in new fields (called "wildcat" drilling) is successful; the other eight attempts result in "dry holes." When we examine the pattern of petroleum production throughout the world, we find that Poland, Mexico, Germany, and several other countries have already passed their years of peak production. These conditions may be partly due to political influences, but even in the favorable production climate of the United States, several of our states have reached their peak production. Such big oil producing states as Oklahoma, Pennsylvania, and West Virginia have passed their peak years, but, fortunately, new fields along the Gulf coast of Texas and Louisiana and in North Dakota have made up for the dwindling production elsewhere, so that our national production of petroleum has continued to increase. The peak of petroleum production in Russia and the Middle East is not expected to be reached until sometime after our peak is passed.

The brightest hope for satisfying our fuel needs in the long run lies in the development of energy from other sources besides fossil fuels. The primary source of energy, the source that supplies the energy contained in our fossil fuels, is solar energy, the origin of which we have already examined. Man has used solar energy in one form or another for centuries, and we need not fear its depletion for billions of years. The use of solar energy directly, however, has not been popular because we have not learned how to use it effectively. The sun's rays may be reflected by a mirror

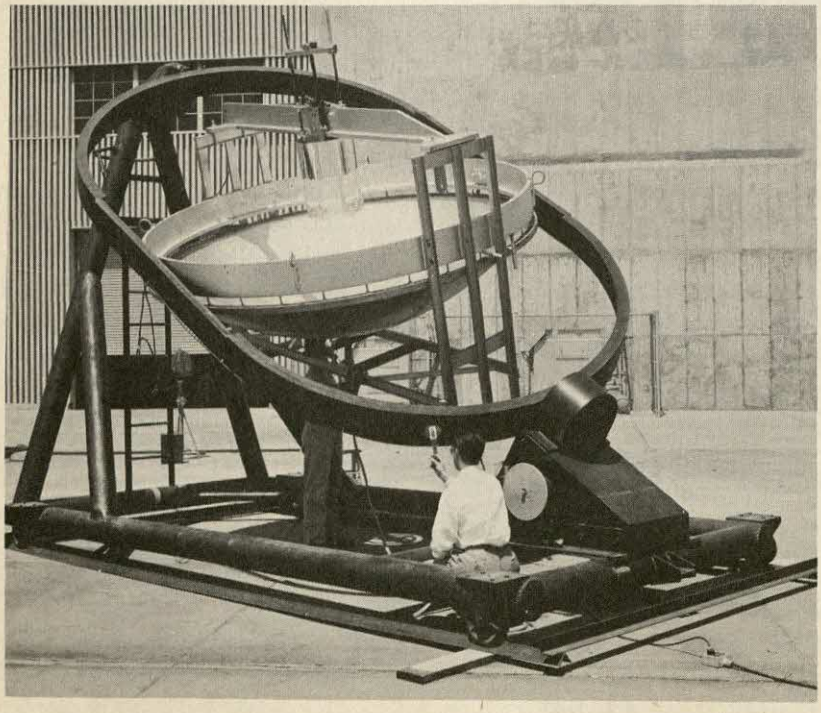
to produce temperatures high enough to cook food. Reflectors, three and a half feet in diameter, are sufficient to bring a quart of water to boiling in January, but if the day is cloudy or if we wish to fry steak at night, solar energy is of little value.

So intriguing are the possibilities of harnessing solar energy for direct use that much research is currently underway on this problem. In the French Pyrenees, a huge 40-foot parabolic mirror is being tested as a practical means of creating high temperatures. This solar furnace can produce 75 kilowatts of power and temperatures up to 3000°C. In the United States, several research groups are using parabolic reflecting mirrors to produce high temperatures. One of these is ten feet in diameter (Fig. 28-1), and others are adapted from the mirrors used in military searchlights.

The conversion of solar energy directly into electrical energy has been achieved by the so-called *solar battery*. A rural telephone system in Georgia has been successfully operated by solar energy stored by use of these cells.

Solar energy is used indirectly in several ways, the most important of which is the storing of solar energy in plants through photosynthesis. In this process, discussed in detail in Chapter 18, plants capture solar energy and store it in the form of chemical energy. This energy may later be released through combustion or through the consumption of

Fig. 28-1. Solar furnace designed primarily for research studies. The sun's rays, focused by a parabolic mirror are capable of producing temperatures up to 5000°C. (Courtesy Convair Division of General Dynamics Corp.)



foods. Our lives depend upon this process, although, on the whole it is not very efficient, and only about 0.15 per cent of the sun's energy arriving at the earth is caught and stored by plants.

Waterpower, which has long turned the wheels of industry for man, is indirectly the consequence of solar energy. Through the action of the sun, water is evaporated from streams, rivers, and oceans and it then falls as rain on the uplands. Part of this water is collected in reservoirs and permitted to flow through hydroelectric generators in a continuing production of usable power from a renewable resource. Many of our best power sites have already been utilized, and silting due to soil erosion is causing increasing concern. In spite of our efforts to employ this source of energy, however, waterpower will never provide much more than 4 per cent of the energy needs of the United States.

Most of the energy of the wind comes from the sun. The interplay of cool and warm air fronts, which produces high and low pressure areas and powerful convection currents, generates considerable amounts of energy. It has been estimated that 2 per cent of the solar energy reaching the earth is converted into winds. Man has never attempted to harness wind energy on a very great scale, although in certain parts of the world wind power has been of vital importance. The Dutch literally pumped the ocean off their land with windmills, but these have largely been replaced with diesel and electric pumps. The windmill on our western plains has also been replaced. In many parts of the world interest in harnessing wind power is increasing, for it is a renewable resource, and yet it will probably never contribute more than 2 per cent to the energy needs of the world.

The tides are another source of energy that might be utilized. The energy of the tides is directly traced to the rotational energy of the earth—in fact, as we have seen, the tides are slowing down the earth's rotation at the rate of about one second every thousand years. This tidal energy is tremendous, but engineers have not yet succeeded in harnessing it to a significant extent. Although there are some small tide-power plants in operation, engineers do not expect this form of energy to supply much more than a trace to the energy needs of the world.

In Chapter 26 we saw how promising the possibilities are of harnessing the energy of nuclear changes. Economically, this source of power is nearly competitive with fossil fuel energy. (It is expected that the cost will be 5.5 to 6 mills per kilowatt-hour at nuclear plants which are presently under construction.) As our energy demands rapidly increase, nuclear energy will become a common means of generating power. Engineers estimate that it may even be the most important source of energy for the world by the year 2000.

Nuclear power holds our hopes for energy in the future, in spite of the problems it brings. People must be made aware of all aspects of this new source of energy: the role

it will play in our society; the importance of developing it in order to lessen the demand on our fossil fuels; the realization that it probably will be in electrical form and what this will mean to industry and society. The problem of radioactive waste disposal must also be solved.

28-2 Conservation of Mechanical Energy

A good many of us, including scientists as well as nonscientists, have probably at one time or another dabbled with the idea of creating energy by a perpetual motion machine. The history of science tells of many attempts, some of which were quite elaborate, to create energy by what might be called "boot-strap" methods. Not only have these attempts failed, they have shown conclusively that creating energy from nothing is an impossibility. In addition to empirical proof, there are sound theoretical principles that rule out such a creation of energy by any process now known to man. Since it is hardly worth our efforts to try to construct a "perpetual motion" machine, let us examine some of the natural phenomena that illustrate the idea of conservation of mechanical energy.

In Chapter 5 we saw that a pendulum in the form of a swing (Fig. 5-12) represents a conservative system. That is, the potential energy of the pendulum is a maximum at the end of the swinging arc, while at the middle of the arc the potential energy is a minimum. The loss of potential energy is accompanied by an increase in kinetic energy, and, as the swing goes to and fro, the potential energy, mgh, is converted into kinetic energy, $\frac{1}{2}mv^2$, and then back again into potential energy. The mathematical equation for this relationship is:

$$mgh = \tfrac{1}{2}mv^2 \qquad (28.1)$$

This equation, derived from Equations 5.12 and 5.13, tells us quantitatively the amount of mechanical energy that is converted continuously from potential energy into kinetic energy in a swinging pendulum.

It is not always easy to account for all the energy changes in a system, but the more we refine our experiments in testing the principle of conservation, the more we are led to the conclusion that energy is always conserved. For example, when we drop a steel ball onto a smooth steel plate, we find that the harder we make the ball and the smoother we make the plate the more nearly the ball bounces to the height from which it was dropped. If we perform this experiment in a vacuum, the results almost substantiate the principle, and if we could reduce the loss of mechanical energy to other forms of energy such as heat, the results would presumably coincide exactly with the conservation principle.

A major railroad that crosses the Rocky Mountains makes an interesting application of the conservation principle,

thereby saving substantial amounts of electrical energy. As a train, pulled by an electric locomotive, climbs a mountain it uses large amounts of electricity, but as it crosses the pass the engineer changes the motors into generators. When the train rolls down the other side, the locomotive generates electrical energy and sends it back into the trolley wire. The kinetic energy of the moving train is transformed into electrical energy, and this slows the train to a safe speed as it descends the mountain. The regenerated electrical energy represents a return of the energy that was required to pull the train up the mountain, and this electrical energy can be used to pull another train up the same grade, or it can be returned to the power company. In either case, the conservation principle is used to reduce the cost of electricity and thus to reduce "fuel" costs.

28-3 Conservation of Energy in Atomic Systems

One of the most amazing principles of physical science was revealed by a study of the energy of the components of atoms. As early as 1902, electrons were discovered to have a mass that varies with their velocity. When high-speed electrons such as beta rays were passed through electrical and magnetic fields and deflected by the well-known principle used in the mass spectrograph, it was found that their mass increases with velocity. We now know that theoretically the mass of an electron or of any particle, charged or neutral, would become infinite if its velocity were to become equal to the velocity of light. This idea was developed by Einstein and appeared in his now famous theory of relativity in 1905. While it is beyond the scope of our work to derive the relationship, the equation that relates the change of mass with velocity is:

$$m = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}} \qquad (28.2)$$

where m_0 is the mass of an object at rest, m its mass when traveling at a velocity v, and c the velocity of light (approximately 3×10^8 m/sec). At ordinary velocities such as those achieved by automobiles or even airplanes, v is so small compared to c that the mass is little affected. But in the study of subatomic particles, we encounter alpha and beta particles that travel with velocities up to 90 per cent of that of light. At these high velocities, the mass of the particles becomes noticeably larger than their mass when they are at rest.

At first thought, this idea of mass changing with velocity may appear strange. Does an object have more molecules in it when it is moving? It is inconceivable that merely imparting velocity to an object would endow it with more molecules. If no tangible matter has been added to the object, then mass cannot be the primary measure of the matter in

an object. How then do we account for the increase in mass of an object when it is moving?

In order to answer this question let us rearrange Equation 28.2 according to an algebraic procedure that need not concern us here. If the velocity of the object is small compared to c, then, through this rearrangement, the equation will become:

$$m = m_0 + \frac{\frac{1}{2} m_0 v^2}{c^2} \qquad (28.3)$$

This expression now answers our question because we can see immediately that the increase in mass is due to the last term in Equation 28.3, which may be defined as:

$$\frac{\frac{1}{2} m_0 v^2}{c^2} = \frac{\text{kinetic energy}}{c^2}$$

Since c, the velocity of light, is constant, the increase in mass of a moving object is directly related to its kinetic energy. This gives us the sum and substance of the principle of the conservation of mass-energy.

Up to now we have dealt with mass and energy as two independent entities. In our ordinary experience with mechanical energy changes, we may continue to do so. However, whenever we wish to account precisely for all the energy in a system, we must realize that when we impart kinetic energy to an object, the mass of that object becomes greater. For each joule of energy supplied to a given mass, that mass becomes greater by 1.1×10^{-17} kg. Since a joule of energy may be expressed as heat or electrical energy as well as mechanical energy, we may make an important generalization. We may say that the total energy of a closed system is constant, that energy may change from one form of energy to another, and that energy may even change to mass. The ramifications of this principle of conservation are far-reaching and interesting. Let us consider a few examples.

Suppose we have 10 kilograms of iron at room temperature. What will happen to its mass if we warm it to cherry-red heat? When we add heat to the iron we are adding energy. We know that one calorie of heat is equivalent to 4.18 joules and that it will require in the neighborhood of 1.5×10^6 calories to bring this piece of iron to red heat. Thus the mass of the iron will be increased by:

$$1.5 \times 10^6 \text{ cal} \times 4.18 \text{ joules/cal} \times 1.1 \times 10^{-17} \text{ kg/joule} = 9.5 \times 10^{-11} \text{ kg}$$

This is a very small, but measurable, increase in the observable mass of the iron.

Another interesting example of the principle of the conservation of mass-energy is shown by comparing the mass of an atom with the masses of its component parts. For instance, the helium nucleus contains four particles: two protons and two neutrons. When these four particles are combined in the nuclear structure, they have a rest mass of

4.0028 atomic mass units. But the masses of two protons (2×1.0076) and two neutrons (2×1.0090) when taken individually add up to 4.0332 atomic mass units, or approximately 0.03 mass units greater than when these four particles are combined in the helium nucleus. It would appear that the sum of the parts is greater than the whole, but we must remember that on account of the binding energy within the nucleus (Chapter 26) a great amount of energy must be supplied to the helium nucleus in order to separate it into its components. In fact, it is the reverse process—the combining of these individual particles—that accounts for the release of the tremendous quantities of energy by the sun. The breaking up of the helium nucleus is endothermic, whereas the building up of the helium nucleus is exothermic. Furthermore, the change in mass accompanying these processes is a direct measure of the amount of energy gained or lost by the system in the change.

A rather startling example of the mass-energy relationship is observed in subatomic studies of the collision between an electron and a positron. Both of these charged particles have measurable masses prior to their collision. However, at the instant of their collision with each other, their total mass is converted into energy in the form of gamma rays, the energy of which is equivalent to the combined masses of the two particles (Fig. 28-2). We may say that mass has been converted into energy, or that mass is a form of energy, or that mass and energy are equivalent to each other, but in whatever manner we say it, the principle, which is well established and proved, is the same.

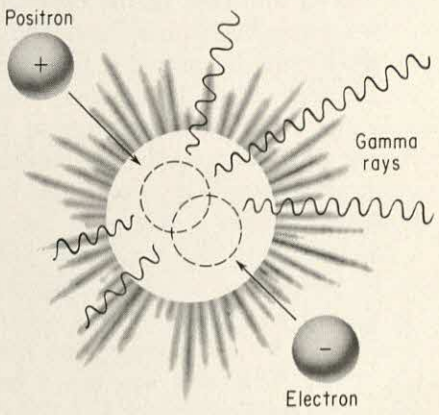


Fig. 28-2. The collision of a positron with an electron in which mass is completely converted into energy.

In an earlier chapter, we encountered the idea of the conversion of mass to energy according to the equation:

$$E = mc^2$$

This equation gives us the relationship between energy and mass when m kilograms of mass are converted to E joules of energy. The quantity c is again the velocity of light. This is merely another revelation of the quantitative relation between mass and energy. The application of this principle to radioactivity, nuclear transformations, fusion, and fission has literally opened the door to the golden age of modern science.

28-4 Conservation of Heat Energy

We have already seen in Chapter 20 that there is a definite relationship between heat energy and mechanical energy. The so-called joule equivalent of heat and mechanical energy is given by the expression: 1 calorie of heat = 4.18 joules of energy, or 1 Btu = 778 ft-lb. In practically all mechanical processes, we encounter friction—that is, a loss of a certain amount of mechanical energy. It is impossible to construct a machine that has frictionless bearings or gears.

Even in operating a well-lubricated automobile, over two-thirds of the energy generated by the engine is dissipated through friction in the drive train, gears, axles, and tires. This "lost" mechanical energy is converted into heat energy, as you may easily ascertain if you examine how warm the tires and the operating parts of your car get after a drive.

Let us study the energy changes that occur in a gas when it is under pressure. When a gas under pressure is released it cools off noticeably. The compressed air that rushes out of a tire, for example, feels delightfully cool even on a hot day. But why does the compressed air cool off when it expands? The answer to this question lies in the conservation principle. The air rushing out of the tire must push back a certain amount of air. That is, the expanding air performs work on the surrounding atmosphere. The energy required to do this work must be provided from some source, and, unless outside heat energy is supplied, this energy is taken from the kinetic and potential energy of the molecules of the expanding air. Consequently, the temperature of the air is reduced in proportion to the loss of this kinetic energy.

The thoughtful person at this point may suggest that we permit a gas to expand into a vacuum. Such an expanding gas needs to perform no work in expanding because it is expanding against nothing. Since it is not necessary to push anything out of its way, its temperature should remain constant (Fig. 28-3). This idea was tested early in the nineteenth century by the Frenchman, L. J. Gay-Lussac, and later by James Joule of England, and found to be true. In other words, if a gas does no work on its surroundings it will not use up any of its internal energy.

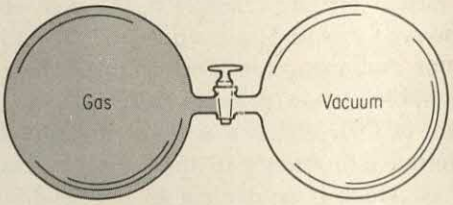


Fig. 28-3. A gas expanding into a vacuum does no work.

The principle of conservation of heat energy is a special case of the general principle of conservation of mass-energy. However, heat energy is of prime importance in practical as well as theoretical matters, for our modern civilization would come almost to a standstill without heat engines of one kind or another. Consequently, a great deal of study has been made on the heat changes that occur in many physical and chemical changes. This field of study is called *thermodynamics*, and it has been extremely helpful to the practicing engineer, the laboratory scientist, and the theoretician.

The law of conservation of heat energy is sometimes called the *first law of thermodynamics*. In simple terms it may be expressed as follows:

heat added to a gas = increase in internal energy + work done by the gas

This law states briefly that heat energy when added to a gas is converted into two other kinds of energy—internal energy and work. Suppose we confine a gas in a cylinder that is fitted with a frictionless piston. If the gas is not permitted to do any work, that is, if it is not permitted to expand, then all the heat energy becomes internal energy (Fig. 28-4). This is understandable if we remember that heat is essen-

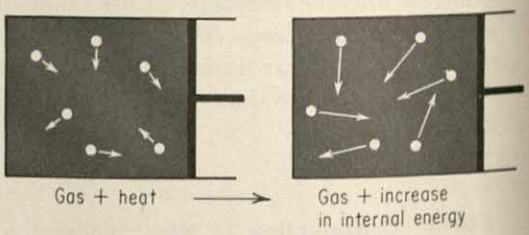


Fig. 28-4. The change in internal energy of a gas heated at constant volume.

tially an energy concept and that the amount of heat present in a gas is manifested by the kinetic and potential energy of the gaseous molecules. But if we permit the gas to do work when the gas is heated, part of the heat energy is used up in doing work (Fig. 28-5). Since work is force times distance, we may measure the work done by the gas if we know the force of the expansion against its surroundings and the distance that the piston moves.

An entirely different example may throw more light on the meaning of the first law of thermodynamics. We have all seen water boil in a teakettle. We know that the water continues to boil in a kettle only as long as heat is supplied to it. Once we turn off the stove, the water ceases to boil. Careful measurements reveal that each gram of water requires 540 cal of heat to convert it to steam at 100°C and at atmospheric pressure. When one gram of water changes to steam it expands 1700 cm^3. That is, the steam from one gram of water must push back the air and expand into a volume of 1700 cm^3, which is like pushing an invisible piston and forcing the atmosphere away to this extent. The work done by the steam as it expands is 170 joules, or 42 cal. Of the 540 cal, these 42 cal are used up in work. The remaining 498 cal become associated with the internal energy of the steam molecules. The principle of conservation of energy tells us here that all the energy added to the water in making it boil is fully accounted for.

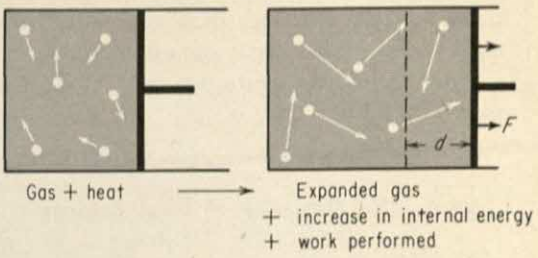


Fig. 28-5. The change in internal energy of a gas that is heated and permitted to do work.

28-5 Heat Engines

There are several kinds of successful heat engines. Piston and turbine steam engines have long been favorites in transportation and for electric power generators. These engines are sometimes called external combustion types. In them a fuel such as coal, oil, or gas is burned and the heat is supplied to a boiler where water is converted into steam at elevated temperatures and pressures (commonly at a temperature of 200°C and a pressure of 200 lb/sq in.). The high-pressure steam is then permitted to expand, either against a piston or through a series of rotating vanes, and it is the energy released by the expansion of this steam that is converted to mechanical energy. As the steam expands, it does work by pushing either the piston or the vanes, and we may calculate the work performed if we know the force and distance through which this action occurs.

The internal combustion engine is used as the motor in our autos, trucks, buses, and diesel trains. In these engines, a combustible fuel, mixed with air, is drawn into a cylinder. Upon the firing of the fuel-air mixture, a rapid combustion occurs that releases the stored-up chemical energy in the form of heat. The combustion products—carbon dioxide, water vapor, and the residual nitrogen from the air—are warmed up suddenly to temperatures of about 3500°F. The

resulting gases at high pressures are then permitted to expand against a piston, thus generating mechanical energy. Again, if we know the force and distance through which the expanding gases operate, we can calculate the work accomplished.

28-6 The Efficiency of Heat Engines

We now see that it is impossible to convert all the energy in a given quantity of heat into mechanical energy. It is inconceivable that we could attain either an infinitely high temperature for the intake or absolute zero for the exhaust temperature. Consequently, we are faced with the fact that no heat engine can operate 100 per cent efficiently.

The efficiency of a heat engine is the ratio of the amount of heat used by the engine to the total amount of heat available to the engine. Thus:

$$\text{efficiency} = \frac{\text{heat added} - \text{heat exhausted}}{\text{heat added}}$$

$$E = \frac{Q_2 - Q_1}{Q_2}$$

Since the heat energy of a system is directly related to its absolute temperature, we may define the efficiency of a heat engine in terms of the temperature of the intake, T_2, and the exhaust, T_1:

$$\text{efficiency} = \frac{T_2 - T_1}{T_2} \qquad (28.4)$$

If a steam engine operates with an intake temperature of 500°K and an exhaust of 300°K, its maximum efficiency would be

$$\text{efficiency} = \frac{500° - 300°}{500°}$$

$$= 0.40 \text{ or } 40 \text{ per cent}$$

Most heat engines operate at much lower efficiencies than the one illustrated above. In an actual heat engine, there are considerable losses of energy through friction. The piston that slides up and down the cylinder, the crank bearing that attaches the piston to the drive shaft, and the system of valves that controls the intake of fuel and air and that releases the exhaust all create friction that detracts from the maximum work produced by an engine.

A typical modern motor car may use only 5 per cent of the energy contained in the original crude petroleum that is the source of heat energy for its heat engine. For instance, let us consider 100 units of crude petroleum as they occur in a natural deposit. It requires 4 of these units of energy to produce the crude petroleum, that is, to drill the oil well and to pump out the petroleum. In the refining process, the

100 units of energy dwindle to 83.5. Transportation to the filling station uses up 2.5 units. The engines of our cars operate at a thermal efficiency of 29 per cent, as measured experimentally. The theoretical thermal efficiency, as predicted by Equation 28.4, would be somewhat higher, but further energy is siphoned off by the auxiliary devices on the car, the drive train, and the tires. By the time the car is being pushed forward against the friction of the wind, only 5 units of the original 100 are doing useful work. Table 28-2 summarizes this information.

TABLE 28-2 EFFICIENCY OF THE MODERN CAR

	Efficiency of each step	*Total efficiency*
Crude petroleum	100	100
Production of crude petroleum	96	96
Refining of petroleum	87	83.5
Transportation of gasoline	97	81
Efficiency of engine	29	23.5
Mechanical efficiency of car	71	16.7
Rolling efficiency	30	5

From Ayres and Scarlott, *Energy Sources—The Wealth of the World* (New York: McGraw-Hill Book Company, 1952).

In spite of the low efficiency of the modern car, the trend is to drive heavier cars that are loaded with more energy-consuming gadgets and equipped with softer tires. At the same time, we insist that our cars must have engines large enough to provide a maximum speed of 100 mi/hr, whether we use this speed or not. Such a car requires an engine that is three and one-half times more powerful than one designed for a maximum speed of 50 mi/hr. The aggregate horsepower rating of registered motor vehicles in the United States is about 4 billion, but because of the low efficiency and the fact that the engine does work only part of the time (not when idling at a stop light, for instance), only 0.2 per cent of this rated horsepower is realized. We are indeed an over-powered people who are dissipating our heritage of fossil liquid fuel at an alarming rate, although we appear to be having a wonderful time doing it!

It is only fair to add that the use of these large quantities of energy provides us with a standard of living that is unequaled in any other large country in the world. Our efficient transportation gives us a unity of understanding that may be one reason why our democracy works as well as it does. Children are transported to better schools; we can travel to see how others live and to share ideas with people thousands of miles from our homes; and we can produce and distribute goods, something that would be impossible without rapid means of transportation.

28-7 The Availability of Useful Energy from Heat

In our examination of heat engines, we have met an important principle that governs the availability of heat energy for useful work: we must operate a heat engine between two different temperatures. We find that heat energy cannot be completely converted into mechanical energy nor can we expect to derive mechanical energy from a single reservoir of heat energy. With this experience in mind, we now come to another basic law of heat energy, the *second law of thermodynamics*. This law is a restriction of the first law, which was essentially a law of conservation of heat energy. The second law may be stated: *heat energy will not flow from a cooler to a warmer body.* Although it would be foolish to try to warm our hands on a cake of ice, there may be considerable heat in the ice. Since the heat in ice is at a much lower level than the heat in our bodies, heat cannot flow from the ice to our hands.

In general, objects never run uphill by themselves. If we wish to have water run uphill, we must supply outside energy to pump it up the hill. The essence of the second law is based on this fact: energy has never been observed to lift itself spontaneously to a higher level. We may see a rock roll downhill but never uphill from rest. A clock gradually runs down because the potential energy of its spring is used up to run the clock; but we never see a clock wind itself—energy in some form must be supplied from the outside to keep it running. Even in the most complex energy transformations, there is a forward direction to the process, for only by superposing an outside energy upon the system can it be reversed. For example, the burning of gasoline in a car creates heat and mechanical energy, but it is inconceivable that the exhaust gases should recombine with the heat and mechanical energy of the process to form the original gasoline. For some strange, yet dependable reason, processes that release energy do not reverse themselves without the expenditure of outside energy upon them.

The spontaneous nature of these energy processes may be described by still another statement of the second law, a statement that will involve two new terms which we shall now consider. One of these terms is *free energy*. Free energy is the driving energy in a process; it is the useful energy that we may derive from any chemical or physical change. In short, the free energy is the net work produced in these processes. In an ordinary dry cell, the chemical change produces both electrical and heat energy, but the only useful, or free, energy that is made available is the electrical energy, since the heat energy is lost as so much "friction." When a pile driver drops its hammer, most of the kinetic energy of the hammer is transferred to the stake, which is forced into the earth. But the hard oak block of wood that the hammer strikes becomes heated to the point of ignition. This heat is lost energy, unavailable for work.

The second new term is *entropy*. Entropy is a somewhat abstract term, yet it concerns all of us in our daily lives. Entropy is randomness. If we hold three coins in our hand and toss them onto a flat surface, what happens? Do the coins remain near each other or do they tend to scatter? If we place three or four white pebbles together on a beach at low tide, will they remain near each other when the tide comes in or will they become scattered up and down the beach? These simple experiments soon lead us to an important concept of natural science—things tend to become scattered and mixed up with other things. Even energy has a tendency to dissipate and become scattered. For instance, the heat energy in a piece of hot iron soon radiates to the surroundings and becomes uniformly distributed in the object and the air nearby. The solar energy that pours out from the sun travels on and on until it eventually becomes scattered throughout the universe. Consequently, we may say that the entropy of the universe (the randomness) is increasing.

Let us now combine the terms, free energy and entropy, into a new statement of the second law; in a thermodynamic system:

$$\text{total heat supplied} = \text{free energy} + \text{entropy}$$

or we may say:

$$\text{total heat supplied} = \text{useful work} + \text{energy rejected}$$

In a simple heat engine, these quantities are related to each other as in Fig. 28-6.

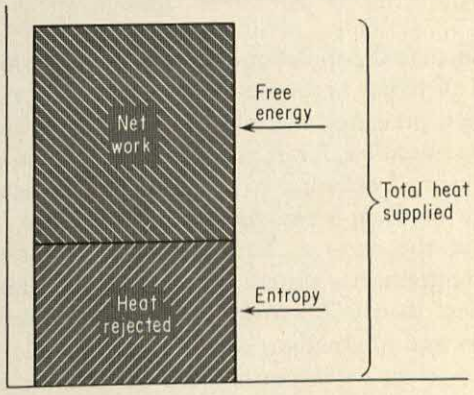


Fig. 28-6. The relationship between heat, free energy, and entropy in a heat engine.

On the college campus, the student body possesses more entropy (randomness) when all the students are at home studying than when they are at a football game. An ordered state of affairs has low entropy. Hence, if all the freshmen sat in one section, the sophomores in another, and so on, there would be less entropy than if all the classes sat together. Any sorting or arranging of the fans in the bleachers would decrease the entropy of the football crowd, and the entropy would be much greater after the game when the spectators dispel and scatter.

In order to reverse the scattering process, outside energy must be supplied. An intentional effort must be made to sort things out, to recover the coins that have been dispersed, to place the white pebbles together again, and to play another football game.

28-8 Energy in Natural Processes

Since free energy is the driving force in an energy transformation and entropy is the scattering effect that naturally accompanies a spontaneous process, the entropy in the universe tends to reach a maximum and the free energy tends to become a minimum. Let us consider several natural processes and find out how these two concepts apply.

That falling objects, when arrested, become warmer was predicted by Joule, who suggested that a stream passing over a waterfall would be slightly warmed by the conversion of potential energy to heat energy. In theory, the temperature of the Niagara River should be warmed about 0.1°C as it passes over its famous 160-ft falls, and a simple calculation will show this to be true. The potential energy is calculated from the equation:

$$\text{P.E.} = mgh$$

Each kilogram of water plunges 160 ft (48.76 m). Thus:

$$\begin{aligned}\text{P.E.} &= 1\text{ kg} \times 9.8\text{ m/sec}^2 \times 48.76\text{ m} \\ &= 477.8\text{ joules} \\ &= 114.3\text{ cal}\end{aligned}$$

Consequently, each kilogram of water is warmed with 114.3 cal of energy, and this accounts for the 0.1°C temperature rise. This calculation has been verified by measurements at the falls, thus further justifying our belief in the law of conservation of energy.

28-9 The Increasing Demands for Our Energy Resources

Increased population and pressures for an increased standard of living both put heavy stress on our energy supplies. These pressures will also be exerted on our soil resources, water supplies, forest products, and mineral resources. Shortages are beginning to occur with increasing frequency in each of these areas. As a simple example, on a worldwide basis, the acres of land available per person for growing basic grains has shrunk from 0.66 acres before World War II to less than 0.55 acres per person at present. Many substitutes and alternatives come to mind to alleviate these shortages—such as increasing use of fertilizer and opening up marginal lands. Almost without exception, however, these alternatives make increasing demands on our energy resources and particularly on our fossil fuels.

These fossil fuels are a nonrenewable resource that reflect millions of years of stored energy from the sun. Unfortunately, convenience has shifted much of our increased demand to those fossil fuels that are in shortest supply. As a result, we have consumed more than half of all the petroleum which has *ever been used* since World War II. (In addition to potential shortages of fossil fuels, the exhausting of so much carbon dioxide and hydrocarbon products into the atmosphere has raised serious problems in some areas, and some scientists think it will make substantial changes in the heat balance of the atmosphere.)

The point here is not that we will run short of energy, but that we will almost certainly change our patterns of energy use as one form of energy or another gains competitive

cost advantage. A recent European Economic Community memorandum estimates that energy costs account for:

25 per cent of the total cost of steel making,
20 per cent of the cost of producing other metals,
10 to 15 per cent of costs in chemical manufacturing, and
8 per cent of the cost in other manufacturing.

The memorandum goes on to state that the cost of energy is one of the major forces determining a country's international competitive position and the location of its industries.

From our study of physical science and particularly this chapter, we should recognize some limits to the efficiency of our machines. We should not be readily misled by a miracle cure, a 200-mile per gallon carburetor, or other similar easy solutions. We should also be able to assess more realistically the coming changes in our energy sources and their meaning to society.

28-10 Summary

Most forms of energy—biochemical, fossil fuel, water power, and wind power—can be traced to solar energy as the original source. Solar energy has only recently been seriously considered as a direct source of energy for our needs.

Nuclear energy, already established as a source of energy for peacetime use, is apparently the only significant form of energy that is not traceable to solar energy.

The principle of conservation of mechanical energy is believed to be valid on the basis of both experimental and theoretical proof. This principle can be extended to include all forms of energy that are involved in a system, whether it be heat, radiant, chemical, or nuclear energy. The expanded principle is commonly called the principle of conservation of mass-energy.

The laws of thermodynamics state that heat energy is conserved in all heat transfer processes, and that energy at one level cannot spontaneously flow to a higher level.

The efficiency of a heat engine is the ratio of the heat energy converted into other energy to the total amount of heat added to the engine.

The free energy of a process is the energy available for useful work, whereas entropy represents the energy that is rejected. Entropy is the energy of randomness or the lack of order.

EXERCISES

A. KEY TERMS TO REMEMBER

efficiency of heat engine
entropy
first law of thermodynamics
fossil fuel
free energy
heat engine
internal energy
law of conservation of mass-energy
second law of thermodynamics

B. QUESTIONS ABOUT THE AVAILABILITY OF ENERGY

1. List all the primary sources of energy that we encounter in our daily lives.

2. Trace the ultimate source of the energy that is expended when we lift a book.

3. What is the source of energy in a so-called self-winding wrist watch?

4. When, in a football game, does the crowd, including the teams (especially when the home team wins), exhibit the most entropy? When the least entropy?

5. In what kind of system may we expect to find zero entropy —that is, perfect order?

6. What happens to the electrical energy of a bolt of lightning after it has struck a tall tree?

7. List the substitutions or adjustments we can make, and for each substitution list the resources and energy sources that will be in greater demand as a result, in each of the following areas where shortages may arise: (a) soil resources, (b) forest resources, and (c) mineral resources.

8. What increases in our standard of living have not come from or resulted in some increase in use of our energy resources?

9. Can you anticipate what the changes are that might arise in our society if the costs of energy (all kinds) were to double?

10. What public policies, if any, tend to encourage the waste of our natural resources?

11. Our energy resources may become such that we will have to rely mostly upon nuclear energy, which will supply the consumer with power in the form of electricity. What changes can you foresee should 80 per cent of our energy be delivered in the electrical form?

C. PROBLEMS ABOUT THE AVAILABILITY OF ENERGY

1. What would be the mass of a rifle bullet that has a rest mass of 10 grams and a velocity of 1000 meters per second?

2. A supersonic airplane with a rest mass of 10,000 kilograms would have to reach what velocity to have a mass twice as large? Hint: Use Equation 28.2. Ans. 2.6×10^8 m/sec.

3. If it requires 10^4 joules of energy to accelerate a typical automobile from rest to a speed of 60 miles per hour, what will be the resulting increase in the mass of the car? Will this increase in mass be observable to the occupants of the car? Assume that the car's mass is 1.5 tons.

4. The annual electrical energy production in the United States is about 4×10^{17} joules. If this were to be generated by nuclear power at 100 per cent efficiency, what would be the amount of matter converted into energy in a year?

Twenty-Nine

New Horizons in Physical Science

In the history of the human race, the last hundred years are but a split second. Yet in this brief interval of time more scientific and technological advancements have been achieved than in the thousands of preceding years. In these few years electricity has been harnessed, the structure of the atom has been unveiled, nuclear energy has been tapped, the universe has been further opened to us, and the welfare of mankind has been improved through the conquering of disease by the development of, among other things, antibiotics, anesthetics, vaccines, and good nutrition. Radio, television, jet airplanes, penicillin, nylon, and atomic power were unknown or were in their embryonic stage a short thirty years ago. Many of the major achievements of modern science and technology have come within our own lifetime.

All these new things cause us to wonder why and how they came so rapidly. But even more wonderful is the ever-increasing rate at which these things are developing. If we were to project into the future the rate at which new discoveries and developments are now being made, we should find that at the end of the next thirty years we may be four or eight or even more times further along than we have come in the past thirty years. An examination of the number of new scientific publications that appear each year in *Chemical Abstracts* gives us an idea of the accelerating rate of growth that is typical of all fields of physical science. In 1910 there were 35,000 scientific articles in the abstracts, in 1930, 156,000 articles, in 1950, 255,000 articles, and nearly a half million in 1960.

Scientific advancements do not usually come suddenly. They grow out of many years of study that often appear

fruitless and unproductive. Occasionally a major advancement comes quickly, but ordinarily the breakthroughs in science come only after many years of preliminary investigations by many workers. Most of the developments listed above were nurtured for years in hundreds of laboratories before they reached fruition. Consequently, by examining the work that is being pursued in the scientific laboratories of today, we may project ourselves into the near future and guess at the new developments that will occur. Before making these predictions, hazardous as they are, let us consider some of the current pioneering research upon which future developments are likely to be based.

29-1 Modern Concepts of Space

When we delve into the modern concepts of space we encounter several mysterious terms, such as relativity, wave mechanics, and the curvature of space. In spite of the apparent uncommon sense that seems to be associated with these new ideas, once we become acquainted with them we will find that they not only make sense but explain certain natural phenomena that cannot be otherwise explained.

A word of warning may be advisable before we launch into an examination of these frontier ideas. We have already seen how the structure of the atom (Chapter 12) requires a much more adequate concept of forces than are described by Newton's laws. That is, the classical laws are not comprehensive enough to fit the smallest parts of the atom. This does not require that we discard Newton's laws as they apply to ordinary objects. It merely points out the necessity to find a more adequate explanation for the atom, and such an explanation is found in the quantum theory. We will also find shortly that Newton's laws do not fit the larger objects in the physical world either—the stars and the spaces of the universe. But we do not discard Newton's laws because they do not exactly fit the behavior of the universe. These newer concepts go beyond Newton's laws and are broad enough to explain the behavior of the atom, of ordinary objects, and of the universe. That is, they include the meaning of the classical laws and at the same time explain the smaller and larger reaches of the physical world.

One of the basic concepts of space is linear distance. We have commonly heard the axiom that a straight line is the shortest distance between two points. But how do we measure the distance between two points? If we use a yardstick for ordinary distances, such as the size of a room or the height of our house, we may obtain a reasonably satisfactory measurement. But what about the distance between New York and San Francisco? If we travel between these two points we may feel that we are traversing a straight line, but actually we are moving along a curved path—the surface of the earth. What appears to us as a straight line is

merely part of a great curve. Even the relatively short distance from one end of a building lot to the other (50 yards) cannot be a straight line—it is approximately one-millionth of the earth's circumference and therefore that fraction of a great circle. We should, then, reconsider how we apply the classical ideas of geometry to the study of lines, distances, and space.

In order to avoid the effect of the earth's surface, we could try to construct a line out in space, where we might conceive of a line as being independent of the curvature of the earth or of that of any other planet or of the stars. But here we encounter the concept, based upon rigorous analysis of space itself, that nowhere in the universe is it possible to construct a perfectly straight line. The reason for this is the basic nature of space itself—space is curved. Hence, any line constructed in space will be curved.

Newton conceived of the universe as empty space populated by celestial bodies. In this empty space there are forces such as gravitation, but otherwise the space itself is empty and does not have the character of an object in the usual sense. The modern concept of space, based upon Einstein's work, visualizes the universe in an entirely different manner. In Chapter 12 and again in Chapter 28 we saw how mass and energy are interrelated. Both experiments and theory have shown that mass can be converted into energy and vice versa. If we think of mass as a concentration of energy, then there is more mass-energy in the form of mass at certain places in the universe than at others. In the earth, the planets, and the stars there is a high concentration of mass-energy in the form of mass. Elsewhere in the universe, where the mass-energy is in the form of radiation or light, the mass-energy is spread out in the form of energy. Consequently, the empty spaces of the universe are not really empty but contain a diluted amount of mass-energy in the form of energy.

The energy that is spread out in the space of the universe is not distributed uniformly. It would be unreasonable to think of the energy in space as being a uniform, continuous stuff when we know that energy in the form of mass is much more concentrated at some places than at others. That is, we would expect to find more energy in the space immediately adjacent to a mass than we would out in the middle of space. We know that there is much more light in the space near the sun than there is in the space at the outer edges of the solar system, and since light is one form of energy we may think of the space near the sun as containing a considerable amount of mass-energy in the form of energy. In addition to the radiant energy that is in the space near the sun, there would be other forms of energy there—gravitational, electrical, and magnetic energy. Consequently, all these various forms of energy create a field around the sun, much as a magnetic field is produced around a magnet (Chapter 9).

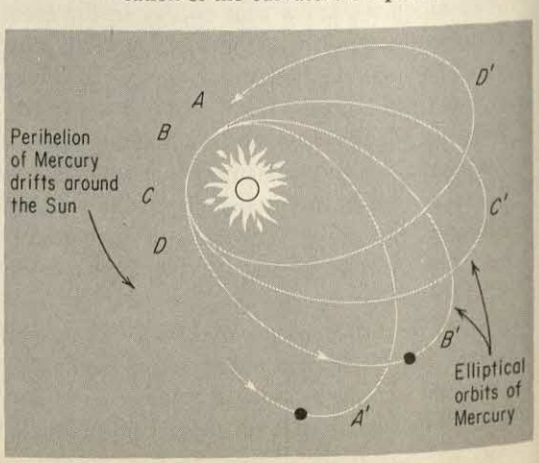
Near any body of mass in the universe there is a field of

energy that is concentrated near the body and that becomes less intense as the distance from the body increases. Since each body is surrounded with its own field of energy, we can expect that among the various bodies of mass that populate the universe there will be overlapping fields of energy, with each body's field extending outward indefinitely from it. These overlapping fields of energy give us our modern concept of space—a region in which varying quantities of energy exist in the form of fields created by the masses of the various objects in the universe. Since this space is not uniformly populated with energy, it will be curved, much as the surface of the earth is curved, where the space is not uniformly populated by mass.

When scientists first heard of this concept of space, there was considerable headshaking. And no wonder, because the idea of space-energy-mass relationships makes uncommon sense indeed. But the support for such an idea was not long in coming. The first proof of the curvature of space came with the solution to a long-standing problem: the irregular behavior of the planet Mercury as it moves in its orbit around the sun. The orbit of Mercury is known to drift in a fashion that cannot be explained by Newton's laws. In a planet's elliptical orbit, the point at which it approaches nearest to the sun (Chapter 7) is called the *perihelion.* The perihelion of all planets except Mercury appears to be at the same point with respect to the sun. That is, these planets go round and round elliptical orbits that remain in the same location. However, the perihelion of Mercury moves gradually in a great circle around the sun once in 220,000 years. Apparently, a force gradually sweeps Mercury away from its normal elliptical behavior that is predicted by Newton's laws of gravitation.

Einstein applied his new concept of space curvature to the problem of Mercury's drifting perihelion and found that Mercury, the planet nearest the sun, is affected by the high concentration of field energy created by the sun in the space near the sun. The planet, in effect, follows the path of least resistance as it moves through the space that is curved. The more curvature in the space (or the greater the field strength in the space) the greater is the planet's deflection from its normal course. A working model of the movement of Mercury appears in Fig. 29-1. At the closest approach of the planet to the sun, the planet is pulled toward the sun more than normally, so that each consecutive revolution of the planet in its elliptical orbit is displaced slightly. The figure shows only four of these revolutions, but after 220,000 years the elliptical orbit will have been pushed around the sun and back to its original position, a phenomenon for which the concept of space curvature gives an adequate explanation. What about the earth and the other planets? Apparently the planets that are farther removed from the sun suffer an unmeasurable effect from the curvature of space caused by the sun.

Fig. 29-1. The perihelion of Mercury as it drifts around the sun—an indication of the curvature of space.



In spite of the successful application of the concept of space curvature to the Mercury problem, much skepticism continued toward the new idea. To counteract these objections, Einstein proposed that his theory of space curvature could be tested directly by the following experiment. If light is a true form of energy, it should be affected by the curvature of space very much as the mass of the planet Mercury is affected. That is, if a ray of light were to pass through space that is highly populated with energy, as occurs near the sun, that ray of light should suffer a change in its normal path. Einstein suggested that the apparent positions of the stars that shine near the sun be examined both before and during a solar eclipse. Ordinarily, the light coming to us from a star should travel on nearly a straight line, but if this light passes near the sun its path should be deflected. The experiment was performed by photographing the stars near the sun during a solar eclipse, and then, leaving the camera undisturbed, photographing the same stars 24 hours later. Supposedly, the starlight during the eclipse should be slightly displaced. These photographs were compared, and once the techniques were perfected, it was proved that the starlight that passes near the sun does suffer a deflection as the result of the concentrated energy field of the sun.

29-2 Relativity

Linked to the whole problem of space curvature is the concept of relativity, which introduces a new idea of time, mass, and space. In Chapter 12 we examined the quantum concept of light as it is related to the structure of the atom. We saw there that an electron in an atom may become excited by an outside source of energy and as a result jump to a higher energy level within the atom. When the electron returns to its normal or ground state, it gives off this energy in the form of light. We found that the color of the emitted light is precisely related to the amount of energy released by the electron. Since each element has its own unique electronic structure, the atoms of a particular element, such as iron, exhibit a unique group of spectral colors when excited—each excited electron contributes to an intricate "fingerprint" of the atom itself.

When we observe the spectrum of iron in our laboratories, the colors have certain wave lengths, but when we observe the spectrum of iron that is in the light from the sun, we find that the colors of all the spectral lines are shifted slightly toward the red end of the spectrum. This means that the light from the iron atoms in the sun has slightly longer wave lengths than does the light generated by excited iron atoms in the laboratory. But why should the wave length of light be different after it travels the 93 million miles from the sun? One answer to this question might be the Doppler effect,

although this explanation is hardly plausible because so far as we know the sun and the earth are not moving toward or away from each other to the extent necessary to produce the red shift. The best explanation of the red shift involves Einstein's concept of space curvature and the energy field of the sun.

According to the so-called "Einstein effect," the movements of electrons in an atom are influenced by the environment of energy (or mass-energy). Where there is little or no energy field, the electrons in atoms move with a maximum velocity; where there is a large energy field, as in the sun, the electrons are slowed down, in which case they will emit lower frequency light when excited. This accounts for the red shift of the light emitted by the atoms that are in the strong energy field of the sun, and from the amount of the red shift we should be able to calculate the mass of the sun. Because there is fairly good agreement between the sun's mass calculated in this manner and the sun's mass obtained by other methods, we may have considerable confidence in the validity of the Einstein effect. The mass of many stars other than the sun's has been determined through an examination of the red shift.

If we apply this concept to ordinary objects, we encounter some strange things. A yardstick at rest has a certain length, but if it is moving it possesses more energy than when it is at rest and, consequently, has more mass (or mass-energy). With a greater mass it has a greater energy field, which in turn causes the electrons in the atoms of the yardstick to slow down, and this produces in effect a shortening of the yardstick. Since the earth moves around the sun at roughly 20 mi/sec, our yardstick will be shortened in length accordingly, and what we call a yard is actually 2×10^{-9} inches shorter than what the yardstick would be if it were standing still.

The diameter of the earth itself is foreshortened in the direction of this movement, so that it is 4 inches less than if the earth were at rest. It is important to note here that we are considering the energy of a moving object as a velocity vector, which means that as the velocity of an object increases, its mass increases and its size decreases only in the direction of the movement. If we could travel in a space ship at 1000 mi/sec, our height and width would remain unchanged but our thickness from front to back would be decreased, and, as we approached the speed of light, our bodies would become as thin as gingerbread men. However, we ourselves would not notice this shrinking effect because the seat on which we sit, and everything in the space ship, would be foreshortened in a similar manner. Since the points of reference change as we change, we would have no way of knowing how much we had shrunk. Similarly, in our ordinary existence on the face of the earth, we do not know how much of this foreshortening occurs because we do not know how fast we are moving, although there is considerable

evidence that our galaxy is rushing rapidly through space. If the earth were suddenly stopped, our yardsticks might conceivably extend to many times their ordinary length, our automobiles might become a block long, and our houses might turn into huge mansions. But all this would happen without advantage to us because we too would extend in size proportionately.

Not only are space and mass relativistically variable, but time itself is variable in these newer concepts. Classically, time was viewed as invariable and uniform; Newton said that time flows equably. Today, however, we know that the flow of time is related to the energy of a system: the greater the energy in a system the slower time will flow. Relativity relationships tell us that a clock would run 5×10^{-7} per cent slower on the sun than it does on the earth, and it is conceivable that our watches run very slightly faster at the top of a mountain than they do at the bottom of a deep mine. Although the difference in the rate of flow of time is imperceptible in our time-keeping instruments, we have good proof that in atoms, which are miniature timing devices, the differences in mass-energy are reflected in the speed of the electrons, and this has, of course, the effect of producing a red shift as noted above.

In summary, the theory of relativity opens for us a new frontier—a frontier that is not very far beyond the bounds of our ordinary existence. As we have seen, there are several phenomena of the physical world that cannot be satisfactorily explained by traditional concepts. We are experiencing here the advancement of scientific theory as it attempts to give us a better explanation for natural phenomena. We are also experiencing an expanding horizon as scientific theory becomes more fully developed. We found that Kepler's laws describe the movement of the earth and planets in the solar system. But Newton's laws give us more understanding because they tell us about the gravitational influence between objects in space. The quantum concepts further extend our range of understanding to the tiny parts of the atom, and, the influences in this realm, we find, are not at all like ordinary gravitational forces. Finally, the Einstein mass-energy law leads us to the broadest concept in all of nature because it embodies all these other laws in addition to areas of human knowledge beyond them, such as the curvature of space. Einstein, in his later years, attempted to develop a unified concept that would relate gravitational and electromagnetic forces and that would thereby further extend our comprehension of natural processes. However, at present, the theory of relativity is the broadest concept that has become generally accepted among scientists.

29-3 New Horizons in Research

The natural curiosity of mankind to learn more about its environment has merely been whetted by the discoveries it

has already made, and the many specialized areas of research that are now underway give us considerable indication that important new developments will be forthcoming. It is always exciting to make new discoveries, and the next best thing is to have a ringside seat when these discoveries are made. With this in mind, let us examine some of these specialized areas of modern research. We will deal with these various areas in an arbitrary order. After all, it is quite difficult to assess the real value of a new research idea until after it has made its impact upon society. For instance, a generation ago a similar list would probably have included a reference to nuclear energy, although it is very improbable that a dependable prediction of its importance to us now could have been made at that time.

In recent years great studies have been made in the basic science of the solid state and electronics. We are all familiar with certain kinds of *semiconductors* and *transistors*, and it can be assumed that much of the new development in these topics will come through engineering and practical applications of these ideas. Nuclear energy and photosynthesis, already quite well established from the standpoint of fundamental science, will also receive continued attention in their development. We will discuss those areas that appear to offer much in the way of new basic science, but no attempt is made to list them in an order of importance or potential usefulness.

Low-temperature research: When metals are cooled down to temperatures near absolute zero, they begin to exhibit a strange property. If a doughnut ring of metallic lead, ordinarily only a fair electrical conductor, is cooled down to near absolute zero, it becomes an exceptionally good conductor. If an electrical current is started around the lead doughnut, the flow of electron seems to move continuously, suffering but slight resistance to flow, and this flow will generate a magnetic field so that, in effect, a piece of lead may serve as a magnet. We call this property *superconductivity*. Other peculiar properties of substances also appear at these low temperatures. One form of liquid helium, for example, refuses to remain in a container, and creeps up the walls, flows down the outer sides, and drips off the bottom of the container. Such a property is called *superflow*. The whole domain of low-temperature research is opening up new frontiers and generating considerable interest among physical scientists. What will these newly discovered properties be good for? At this time we cannot say.

High-temperature research: Without a doubt, the highest temperature man has produced is in a thermonuclear bomb. However, other methods of achieving high temperature are being explored, and one of these warrants a brief description. In Chapter 9 we saw that moving electrons may generate a magnetic field, and, vice versa, that a magnetic field may induce an electrical current. We saw further that there is a definite relationship—the left-hand rule—between the

magnetic poles and the direction of the electrical current. Now if a surge of charged particles, such as electrons, protons, or deuterons, is forced quickly into a strong magnetic field in the direction opposite to that stipulated by the left-hand rule, the particles will be abruptly stopped and forced back the way they came. Consequently, the kinetic energy of the moving particles is converted into heat energy, and temperatures up to 700,000°K have been achieved.

Long-range forces: In the domain of colloids and surface phenomena, where distances of ten to a thousand Angstroms are concerned, there are forces that have not yet been specifically defined. At smaller distances the chemical valence bonds and association bonds (sometimes called van der Waals bonds) primarily involve intermolecular forces. And at greater distances the forces are described by Newton's laws. In these intermediate distances, the forces apparently decrease as the inverse cube of the distance, obviously making them different from those that decrease as the inverse square of the distance. Many important phenomena involve these so-called *long-range forces*—catalysts, the structure of polymers, the behavior of proteins, the mechanism of biological processes, and the action of antigens on bacteria. Here is a broad frontier for the research scientist that extends into many fields of study. With a more adequate understanding of long-range forces will probably come an explanation for many problems that now confront physical science.

Proteins, enzymes, and cell structure: In the areas of biophysical research, probably the most challenging frontier is that of the cell. A cell is a tiny functional unit of a biological system such as a plant's or animal's. It consists of a nucleus of protein substances, of cellular fluids, and of a wall or membrane; many are large enough to study under a microscope. The current interest in cell structure is to gain a better understanding of the chemical processes that occur in the nucleus, which takes in raw materials, transforms energy, regenerates proteins, and metabolizes carbohydrates and fats. Most cells, to continue their kind, split in two and then grow normally, but occasionally cells grow wild and rampant, with the result that cancer and other malignancies develop. With more information about the cell processes may come cures for these malignant diseases. Studies of the cell, the proteins that make up cells, and the complex but important catalytic substances called enzymes are bound to be of utmost importance to the welfare of mankind.

Chemical synthesis: Although several hundred thousand different chemical substances are now known and most of them have been synthesized, this is only a start. The potentialities are almost limitless, since more and more new chemical compounds are being produced every day and there is no reason why millions and millions of different substances cannot eventually be made. With the creation of a new substance comes the thrill of examining its properties, and some

of the new substances almost always find their way into useful end products. We have already examined a few of the outstandingly successful synthetic substances such as plastics, fibers, drugs, and dyes. The surface has only been touched, and research in the synthesis of new chemical compounds will almost surely produce even more exciting and useful materials than those we now have.

The upper atmosphere and space sciences: There are but few geographical frontiers that have not been explored. The atmosphere, the oceans, and the crust of the earth are three areas that still challenge the efforts of the scientific investigator. Spurred on by military interest in long-range rockets, scientists have turned their attention to the upper atmosphere in studies that promise interesting results. We know from our observations of the aurora borealis, meteorites, and the reflection of radio waves that the atmosphere extends upward to certain limits. As the atmosphere thins out, the temperature undergoes peculiar changes, and there are regions of unusual behavior in it. Recent research of the upper atmosphere is concerned with the encounter of the earth with meteorites or meteoric dust. There is a worldwide tendency for extraordinarily heavy rainfalls to take place at certain periods of the year, and one of these periods is on January 12–13. Supposedly the earth at this time passes through a swarm of meteors or meteoric dust, and the dust is picked up by our atmosphere and made available to form raindrops. A raindrop, you will recall, must have a nucleus on which to begin its growth, and dust particles are probably the most common particles that fill this requirement.

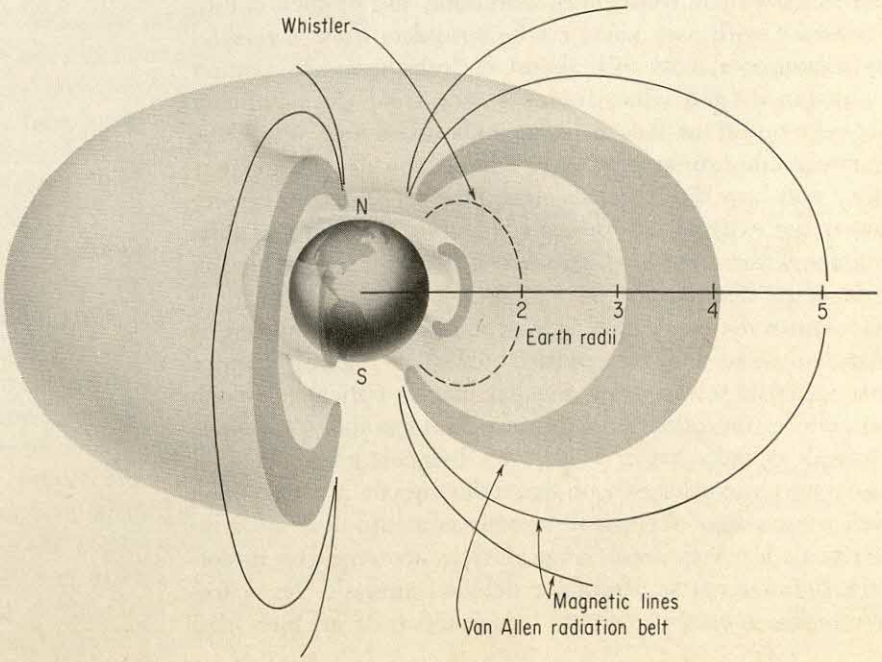
Along with these studies of the outside sources of dust that may account for unusually heavy rainfall have gone attempts to induce rain by artificially seeding the clouds with various kinds of dusts. These efforts have met with only partial success, but the information and techniques that are obtained from them may eventually enable man to increase his control of the weather.

One peculiar characteristic of the atmosphere is its upward and downward oscillation. It has been known for some time that twice daily there is a maximum in the barometric pressure, at 10 a.m. and at 10 p.m. The atmosphere bounces up and down as though it were a giant spring, and the force that keeps it bouncing is the sun. The rotation of the earth exposes the atmosphere to the alternating gravitational forces that cause the tides in the oceans. The gravitational force of the moon is chiefly responsible for the tides in the ocean, but it is the gravitational force of the sun that appears in the proper frequency to match the frequency of the bouncing atmosphere. We have already considered the idea that the moon's action on the ocean tides serves as a brake on the earth's speed of rotation, but there is also the possibility that the sun, as it rises and sets daily, matches the periods of oscillations of the atmosphere and supplies a twist or push to the atmosphere that tends to speed up the rotation of

the earth. The strength of this twist or push has been calculated, and it is possible that the strength of this twist is comparable to the braking strength of the ocean tides. If this is true, the earth's speed of rotation could be speeded up or slowed down, depending upon the relative magnitude of these two forces. We have some evidence to indicate that the earth at one time rotated more rapidly than it now does.

The structure of space: Most of the glamorous aspects of space news deal with the engineering feats of space travel. But with many successful space launchings and with most of these still in orbital operation, we are gaining considerable basic information about the structure of space. The Venus probe and moon landing projects must be considerably refined in order to find much more than is already known through conventional astronomy. Space vehicles have, however, added greatly to our knowledge of the upper atmosphere and near space. The discovery of a layer of ionized particles at heights upward of 16,000 km above the earth's surface at the equator, the Van Allen belt, has led to several tentative hypotheses. One, for example, is the supposition that the aurora borealis is the result of interaction of the earth's magnetic field with the Van Allen belt. The interaction of plasma (ion streams) from the sun with the Van Allen belt is another possible cause of the aurora borealis. Radio signals, called *whistlers*, have now been attributed to lightning at the earth's surface, and these signals follow paths parallel to the lines of force in the earth's magnetic field (Fig. 29-2). Whistlers detected at a point on the earth's

Fig. 29-2. Magnetic lines about the earth and regions of trapped ions known as the Van Allen radiation belt.



surface provide a clue to the location of a lightning storm possibly halfway around the world.

One of the enigmas of the solar system is the abundance of water on Earth but very little water, apparently, on Mars and Venus. The molecular velocity of hydrogen is sufficient to permit it to leak off the earth and escape into space, but at 30,000 to 60,000 ft above the earth's surface it is suggested that molecular hydrogen combines with oxygen to form water vapor. At these elevations, this vapor forms a special type of cloud called "Mother of Pearl" clouds. It is unlikely that surface water vapor reaches these heights. On Mars and Venus little or no such capture of molecular hydrogen is made, and consequently these planets are apparently without a significant amount of water.

Solar energy and photosynthesis: Man's continuing interest in finding new sources of energy and in making more efficient uses of currently available sources will impel him to study in more detail the energy from the sun. Solar energy is one of the few sources of energy that is readily available to almost everyone, yet it comes to us at such low levels that an efficient means of capturing and using it must be found. Already solar cookstoves, steam-electric generators, and solar batteries have been devised to run on solar energy, but this is only the beginning. With a better understanding of photosynthesis may come even more promising methods of using solar energy. One way that has been suggested to utilize directly this energy is called a "solar catalytic cracker." With the proper catalyst, solar energy might conceivably convert the 0.035 per cent of carbon dioxide in the air, along with water vapor, to carbon monoxide and hydrogen. These two gases could constitute the raw materials for a host of synthesis reactions that are already being widely used in the chemical and oil industries.

Astronomy: The frontiers of astronomy are of such nature as to strike with awe anyone who considers them seriously. The astronomer must rely almost entirely upon observation at a distance (and what distances they are!). In astronomy today the questions that are commonly asked are: Why does a star explode into a supernova, what is the structure of our galaxy and how did it get its start, how are galaxies formed, what is the explanation for the red shift, and what scientific evidence is there from which we can derive a theory for the origin of matter itself?

Radioastronomy: A new way of exploring the universe is through the radio waves emitted by certain regions of outer space. A radio telescope, somewhat like an optical telescope that collects the radiation by means of a parabolic reflector, is aimed at those parts of the sky being studied. The intensity and wave lengths of the radio signals are examined much as the light from distant sources is studied. Because the signals are very faint, huge devices are required to collect a radio signal sufficient for detailed analysis. The parabolic mirror must be accurate to a fraction of an inch, and

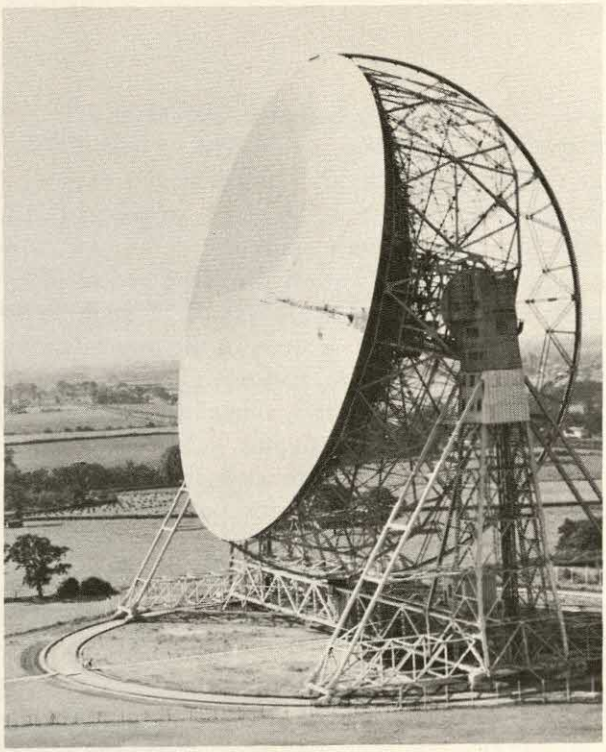


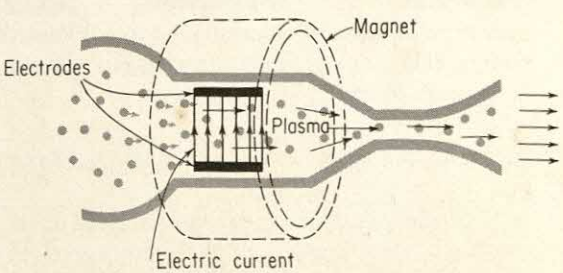
Fig. 29-3. The radio telescope at Jodrell Bank, England. This instrument sent the first messages reflected from the moon and Venus.

changes in temperature and the effects of winds constitute a challenge to the engineer designing such a device. The limit for a single mirror has probably been reached with the 250-foot instrument at Jodrell Bank, England, Fig. 29-3, and attempts to build larger ones, such as the 600-foot telescope in West Virginia, have met with almost insurmountable difficulties.

Several thousand sources of astronomical radio signals have been catalogued and about one hundred of these are directly associated with specific galaxies or parts of galaxies outside our own galaxy. Some of these signals come from intense, concentrated sources, whereas others come from diffuse or regional sources; in addition there are sources of radio signals within our galaxy. Tentative explanations for the cause of these signals have been proposed. For example, it is suggested that the collision of cosmic rays with interstellar gas altered by the magnetic field of a galaxy may produce radio signals. But with more study of these signals we may anticipate that a better understanding of their source and their relationship to galactic structures will be achieved.

Plasma: The characteristics of plasma, a stream of gaseous ions in a vacuum, are being exploited with the objectives of learning about the constitution of matter and the possibility of making a motor that will power a vehicle in outer space. The principle of the plasma motor or ion engine is shown in Fig. 29-4 where gaseous ions evolved from a heated source such as cesium are passed through an electrical and magnetic field. We saw in Chapter 9 how the combination of electrical and magnetic fields exert forces that drive an electric motor, and in this plasma motor the cesium ions are accelerated sufficiently to create a thrust as they are ejected from the

Fig. 29-4. A plasma motor. In the plasma motor an ionized gas passing through an electrical and magnetic field is accelerated and then discharged through a nozzle.



nozzle. Such a device, if successfully designed, should drive a space vehicle at speeds approaching that of light. It would not, however, have sufficient thrust to lift a rocket away from the earth's gravitational field, but would be ample to propel a vehicle once it reached interplanetary space.

Oceanography: A new horizon of research in the physical sciences that combines many of the fundamentals of geology, physics, chemistry, and biology is a study of the oceans. Very little is known of the environmental factors in the ocean that support various forms of plant and animal life. The charting of the temperatures, currents, depths, salinity and mineral content of the oceans is a long-range task. Oceanographers are also studying the relationships between the lowest forms of life, the planktons, and the higher forms of life that are supported by the lower forms. The oceans are a vast storehouse of minerals (magnesium is already commercially obtained from the sea) and it contains tremendous amounts of foodstuffs. Only a small fraction of the resources of the ocean have been developed, and as the world's population increases we will be forced to turn more and more to this part of the earth as a source of raw materials to support mankind. Fairly good progress is already being made in making ocean water potable, and this process promises to supply ample fresh water to regions bordering the ocean that otherwise are too arid to support farms and cities. It has been estimated that certain areas of the ocean's floor contain upward of ten million dollars worth of manganese and iron nodules per square mile. The magnitude of the oceans—3.3×10^8 *cubic miles* of sea water—and the huge amounts of substances dissolved in it—a cubic mile of sea water contains over 1.6×10^6 *tons* of salt—presents a challenging frontier of research.

In summary, we see that the horizons of physical science are constantly being pushed back, and that the rate of discovery of new ideas, concepts, principles, and facts is being accelerated. We live indeed in an age of exciting scientific explorations, and many new discoveries seemingly lie at the immediate edge of the expanding horizon.

EXERCISES

A. KEY TERMS TO REMEMBER

cosmic rays	oceanography	superconductors
curvature of space	plasma	Van Allen radiation belt
energy field	radioastronomy	whistlers
long-range forces	relativity	

B. QUESTIONS ABOUT NEW HORIZONS IN PHYSICAL SCIENCE

1. If the principle of the curvature of lines on the surface of the earth were applied to the construction of a large building, would the opposite walls of the building be exactly parallel to

each other? Would the architect be required to take into account this principle if the building were 100 yards wide? Explain your answer.

2. Is it safe to presume that two million dollars invested in scientific research would produce twice as many results as the investment of one million dollars?

3. The atmosphere of the earth is a protective envelope without which life on the earth might be entirely different from that we know. Make a list of the various things that come into the earth from outer space but are slowed down or absorbed by the atmosphere.

4. Collect from a current newspaper or newsmagazine several articles of scientific interest that bear upon the topics mentioned in this chapter. Using these clippings as a starting point, write a short essay on "A New Horizon in Physical Science."

5. Read an article in one of the issues of the past year from one of these journals, and write a 200 word summary on it:

Scientific American
American Scientist
Science (AAAS)
Science News Letter
Physics Today
Journal of Chemical Education
American Journal of Physics

Thirty

The Philosophy of Physical Science

Most scientists pursue their work with little or no thought about the philosophical aspects of their specialty. That is, the typical scientist studies the physical world through his senses, believing that his observations reflect the existence of a real world. But to the philosopher such an attitude is rather too simple; he takes a more critical view toward facts and the principle of reality. The philosopher raises the questions of what these observations mean and of how the theories that the scientist develops fit into the general structure of knowledge, the nature of physical laws, and the meaning of knowledge itself.

Philosophy, literally speaking, is the love of wisdom. It concerns itself with the study of knowledge. There is much debate among philosophers about the interpretation of knowledge, and such debate is to be expected because many philosophical concepts cannot be tested by experiment in the way the concepts of physical science can. However, it will be profitable for us to consider briefly some of these philosophical ideas in order to gain some appreciation of the relation of philosophy to science.

The common man usually views the external world of the universe as reality. He looks upon the universe and all it contains as a world that exists whether we interpret it or not. This view of the world is called "Realism," and we commonly abide by it in our daily affairs. If in our travels we become lost and inquire the directions of a native, we usually follow his directions without question. Even though we have never seen the road ahead, when the native describes

all the turns and landmarks, we accept for a fact the existence of such a road. As the result of many similar experiences, we have learned to accept for reality many things which we cannot see or have not yet seen. The scientist also relies heavily upon the concept of realism, and, like the lost traveler, he does not hesitate to project himself into the unknown. No one has seen the inside of the earth or the interior parts of an atom—yet we cannot refute the existence of such domains.

In opposition to the concept of Realism are many philosophical ideas, only two of which we shall pause to mention here. One, "Idealism," holds that what we see and perceive are not realities, but that the perceptions themselves are the only realities. The idealist believes that only our *concept* of a meter is real and that when we apply this concept to a stick we have a meter stick. Another viewpoint, "Positivism," holds that our *concept* of a meter is irrelevant, for a positivist believes only what he can experience through his senses or what he can deduce from sense data. Even though most scientists would agree with positivism's meaning for scientific work in general, there are few scientists who are actively interested in the philosophical arguments. Einstein has said that a scientist may "appear as realist, insofar as he seeks to describe the world independent of the act of perception; as idealist insofar as he looks upon concepts and theories as the free inventions of the human spirit; and as positivist insofar as he considers his concepts and theories justified only to the extent to which they furnish a logical representation of relations among sense experiences." Let us now examine in a philosophical manner some of the factors with which the scientist deals in physical science.

30-1 Observation in Physical Science

We gain knowledge through two kinds of experience—direct and indirect. Fortunately, much of the knowledge and information that we have has been passed down from generation to generation by word of mouth, in books, and in various records, thus freeing us from the necessity of gathering it directly. For most of the knowledge in this book on physical science we have relied upon indirect information—it would require several lifetimes to conduct all the experiments and observations that are involved in any one chapter. To re-perform all these experiments in order to achieve firsthand experience would be exciting, but no new goal would be achieved. It is more fruitful to accept (with the usual reservations) what has already been found and to push ahead into new experiments and observations.

Historians of science usually refer to Roger Bacon of the thirteenth century as one of the first to recognize the importance of observations when he said that "without experience nothing can be sufficiently known; he who wishes to

enjoy without doubt the truth of things, should know how to devote his time to experiment." Some years later the famous artist-architect-engineer-scientist Leonardo da Vinci said: "All sciences are in vain and full of errors which do not terminate in observation . . . through one of the five senses." And Galileo was convinced that the solution of natural problems begins "as sensible experiments and necessary demonstrations."

Observation, coupled with reasoning, is one of the most powerful methods of solving problems in physical science. For example, let us follow Newton's observations in his study of light. He had just passed sunlight through a prism and seen the beautiful colors of the rainbow flash upon the wall of his laboratory. Since these colors came either from the sunlight, the prism, or a combination of the two, he reasoned that sunlight is a mixture of colored rays with differing refraction characteristics and that it was this difference in refractability that caused the white light to be separated into its colors. Next, Newton set up several experiments to test his ideas. When he passed blue light through the prism he found that it was refracted more than red light. Then he "mixed" the various colors that he saw on the wall of his laboratory and found that they re-formed white light. From this he deduced that a red object, such as a rose, appears red in sunlight because it reflects only the red rays and absorbs the others. He went on to verify this idea by an ingenious experiment. This brief account of Newton's experiments and reasoning is but a small sample of the imaginative research detailed in his famous report, *Opticks*.

Observation without critical analysis of the experiment may lead to erroneous conclusions. Whether in science or in our ordinary daily lives, our conclusions should be based on information that is collected objectively. This requires that the observer be impersonal, unprejudiced, and without preconceived concepts about the phenomenon he is observing. In his openmindedness he must be willing to accept the data that his experiments produce and at the same time be wary of the data that present themselves. For instance, he may be examining an optical illusion and may observe something that is not in reality there. The thousands of reports of flying saucers in recent years indicate the danger in making observations without a critical viewpoint. Most of these observations of flying saucers have been definitely attributed to balloons, weather conditions, meteorites, and other ordinary natural phenomena.

Once the scientist obtains the observations he is seeking, he must analyze them. This is the ultimate aim of science, for it is the analysis of observations that leads to a systematic ordering of experience. Through analysis, the scientist integrates or sums up all his data under a general law which is, ideally, simple in concept. From this law it is then possible to predict future events. Let us now examine the meaning of some of the analytical operations used in science.

30-2 Cause and Effect

In any particular situation, the idea of cause and effect may be analyzed quite simply. We can say that the cause always precedes the effect in point of time, and we can also say that this particular effect never occurs unless there is a particular cause. If, as the result of many experiments, we note that event *A* always precedes event *B* and that event *B* never occurs until after event *A* takes place, then we say that event *A* is the cause of event *B*. This is a useful and dependable concept that we employ commonly in our everyday activities. If we throw a ball into the air, the ball falls to the earth. We say that every time a ball is tossed into the air, it falls. This relationship is so dependable that a game of baseball can be played with the certainty that any ball goes up must eventually come down. There is no need for the rule book to so specify even though the batter may wish he could strike a ball so hard that it would never come back to earth.

But upon closer inspection of the ball that is tossed into the air, is it necessarily the action of tossing that causes it to fall to earth? Or is it the force of gravity? Or again, is it the force of gravity or is it the earth possessing a gravitational field that is the true cause of the falling ball? In making a critical analysis of the cause and the effect here we soon encounter difficulties and find that the principle of cause and effect is not nearly as precise as we might think. We have all expressed this feeling in the classic question of which came first, the hen or the egg?

30-3 Probability and Hypothesis in Physical Science

The concept of probability in science is quite different from our ordinary idea of probability. We sometimes say: "What is the probability of rain tomorrow?" or, "It is probably cooler today than it was yesterday." In its scientific meaning the concept of probability concerns a phenomenon in which there is a great number of repeating events. If we toss a coin into the air a thousand times, we may examine the probability of the coin landing "heads" or "tails" a certain number of times. The movement of molecules in a gas, the mass of a hydrogen atom, the life expectancy in life insurance schedules, and the half-life period of a radiactive substance are situations in which the concept of probability may be applied with a high degree of accuracy.

The concept of probability may deal with the principle of randomness. For example, if from a million lottery tickets, ticket No. 500,000 were selected as the lucky one, the newspapers would make quite a play on the "improbability" of the selection of this particular number. The headlines would likely read, "Mr. X Reaps Dividends from One-in-Million Chance." The laws of chance would receive considerable

public attention because of the selection of this particular number, but if the selection were an arbitrary one there is exactly the same probability of ticket No. 500,000 being selected as, say, ticket No. 734,619, or any other of the million tickets. The probability of selecting any one of the million numbers is one in a million, and since No. 500,000 is one of the million, *it* stands a chance equal to that of any other ticket in the lottery.

This leads us to a general definition of probability: The numerical value of probability is the ratio of the number of ways an event actually occurs divided by the total number of ways in which that event might occur. In the lottery ticket selection only one ticket is actually drawn, but a million different tickets might be drawn, hence the probability is one in a million.

Once a mathematical relationship can be established between an event and the probability of the recurrence of that event, the scientist is in a position to hypothesize, or guess tentatively about an explanation for a particular phenomenon. For instance, if a scientist is determining the atomic weight of silver, he could not examine, or even sample, all the silver in the world. Instead, he selects what he believes to be a representative sample of silver and, with a very high probability that the silver atoms in his sample have atomic weights similar to all silver atoms, he analyzes only a small fraction of the silver in the world. Again, if a study is being made of the mass of the hydrogen atom, the mass of a large number of hydrogen atoms is determined. With a high probability that each hydrogen atom is identical to the next one, we can arrive at an average, or probable, value for the mass of a single hydrogen atom. Probable values are known for various other physical constants, including the charge on an electron, the electrical equivalent of heat, the velocity of light, Avogadro's number, and so on.

If a probability relationship can be established, a reasonably sound hypothesis may be drawn. Such an hypothesis may appear to transgress the limits of the probability in an experiment, yet it is through a knowledge of probability that the scientist is capable of seeing beyond the limits of his data and finding a fundamental hypothesis, or explanation. In the early study of gravitation, Galileo observed that a heavy object and a light object fell to earth in *almost* the same time interval. According to the actual data of the experiment, he should have concluded that heavy objects always fall somewhat more rapidly than lighter objects. But Galileo perceived that beyond his data there was a constancy, a high probability that a uniform field was acting on the falling objects, so he disregarded the imperfections in his data and arrived at the concept of gravitation. It is not always wise, however, to disregard the slight imperfections in one's experimental data, for, you will recall, the slight deviation in the orbit of the planet Uranus led directly to the discovery of the planet Neptune.

The importance of probability in formulating an hypothesis can be appreciated if we turn to the early studies on atomic structure. Rutherford, in bombarding gold atoms with alpha particles, found that only rarely did the gold atoms alter the path of the alpha particles. Only one in several million alpha particles seemed to meet any sort of obstacle as they passed through the gold atoms. From this low probability of alpha particle deflection, Rutherford reasoned that the gold atom must consist mostly of space and that only a small fraction of the atom had sufficient mass and charge to deflect alpha particles. From the actual, numerical value of the probability of deflection, the size of the atom's nucleus was calculated. Later and more elaborate experiments have verified Rutherford's concept, so that it appears to be a dependable hypothesis, even though based on a consideration of probabilities.

30-4 Scientific Law

By this time we have come far enough in our study to note the role of scientific law in the development of physical science. We have seen that a law is usually a simple summary of a mass of facts concerning a phenomenon of nature. The law of gravitation, for instance, describes the force of gravitation between the earth and all the objects on the earth's surface, between the planets, and, in general between objects anywhere in the universe. Even though a law may originally be based on a limited number of facts, once it is formulated there is an eagerness to test it and apply it to new situations. We know that a scientific law is not to be confused with a political, social, or economic law because the latter are results of legislative, social, or group decisions and are usually based on value judgments. A scientific law, on the other hand, is an interpretation of a natural process, and, although it is humanly conceived, it is based on observations of nature.

Since, as we have seen, experimental observations and measurements cannot be made with perfect accuracy, the question is raised: How accurate can a law be if it is based on data that are only approximate? In other words, what is the true meaning of a scientific law? The classical laws, such as Newton's laws of motion, are based not on observational data, but rather on the notion that behind the data there is coherence and, if you will, the reality of nature. Each experimental observation gives us a glimpse, more or less accurate, of this reality. We have already seen how Galileo in his study of gravitation looked beyond his crude and inaccurate experiments and discovered a remarkably simple truth of nature—the principle of gravitation. Let us consider this matter in more detail concerning a law in another area of physical science.

The gas laws deal with the behavior of gases when they

are subjected to various conditions. We saw in Chapter 20 that the volume of a gas increases with increasing temperature and that the volume of a gas decreases with increasing pressure. These relationships are summarized by Charles' law and Boyle's law, respectively. But these gas laws are commonly referred to as "ideal gas" laws, that is, they deal with what is called the "ideal case." Actually, no gas strictly obeys these laws. If we were to carry out an experiment on the volume-pressure relationship of a real gas such as oxygen, we would find that, especially at high pressures, the behavior of the gas would deviate considerably from Boyle's law. The deviations may be enough to cause us to wonder if Boyle's law is of any value as a scientific law. But we must recall that laws are based on the concept that there is a coherence in nature, and that even though there may be deviations in the facts, a law is judged by whether or not it represents a fundamental relationship. We further test a law by analyzing its meaning—if it is a reasonable and logical argument for the uniformity of nature, then it merits our attention.

With Boyle's law we find that many real gases behave nearly as predicted, particularly at pressures less than one or two atmospheres and at relatively high temperatures. Furthermore, Boyle's law is reasonable and logical, and it is not difficult to imagine an "ideal gas" that behaves exactly as predicted by this law. This places us in a dilemma—shall we discard Boyle's law because real gases cannot be fitted to it, or shall we look for an explanation for the misbehavior of real gases? Both recourses have been pursued, but the latter alternative will concern us here.

If a real gas disobeys Boyle's law, we must find a suitable explanation for the deviation. Since these deviations become most severe at high pressures, it becomes apparent that gas molecules do not exist as "point" objects but actually occupy an appreciable volume. This sounds reasonable, so the next step is to make an allowance for the actual volume of a gas that is due to the molecules themselves. When we make this adjustment in our calculations, the gas appears to obey the law more nearly than before. But there are still deviations, so we look further for possible explanations. One explanation involves the gravitational forces between molecules. At low pressures the molecules are, on the average, far enough apart so that they exert very little intermolecular attraction, but at higher pressure these intermolecular forces may become quite large. Hence, we make another adjustment in our calculations by providing for the intermolecular forces. After making these two adjustments, we find a surprisingly good agreement between the behavior of the real gas and the statement of the ideal gas law.

The significant idea in this discussion is that a scientific law is a description of a natural uniformity. Although it may not be possible to illustrate accurately this uniformity with an experiment, the uniformity does exist, and as we

perfect our experiments and take into account more and more of the causes of the inaccuracies in them we may approach closer and closer to the "ideal case." In the meantime, the law serves the useful purpose of providing us with a guidepost and marker in our pursuit of scientific knowledge.

30-5 Social Attitudes toward Physical Science

Prior to 1600 it was dangerous to be called a "scientist." Although many pursued scientific studies, their primary vocation usually concerned something else. For example, Robert Boyle was a writer of moral and religious essays, but today it is his studies of the behavior of gases that have made him famous. Even in ancient Greece there was a mild prejudice against scientific pursuits, particularly toward applying science to practical situations. Certain phases of mathematics, such as geometry, were well developed long before the birth of Christ, yet there was a reluctance to apply geometric principles in land measurement. Plato felt that such an application would vulgarize mathematics. Archimedes apologized for his mechanical devices, claiming that they were merely gadgets for his own amusement. And later, Galileo, Copernicus, and Tycho met with similar opposition from society, no matter how useful or valuable their ideas were.

In recent times, a slightly more cordial attitude has been shown by society toward science and invention. The change from hostility to friendliness came slowly, however. Early attempts to build steamboats met with disastrous results. In 1707, a steamboat built by Dennis Papin was destroyed by an angry mob of seamen, and Papin narrowly escaped with his life. In 1700 and again in 1760, power-driven sawmills were built in England, only to be immediately demolished by hand-sawyers. But in America, a pioneering land where manpower was at a greater premium than in England, power-driven machinery gained an earlier foothold, and by 1800, boats and sawmills were commonly being powered by steam engines and waterpower, respectively. But the general acceptance of new ideas was slow in coming. In 1805, John Stevens proposed a tunnel under the Hudson River, and after being ridiculed for his idea, Stevens ruefully commented about the "prejudices that naturally arise against anything bearing the stamp of novelty." The Hudson tunnel was built seventy years later.

In pure science, there was also a cool reception to new ideas. Helmholtz in Germany was at first refused publication of his paper on the conservation of energy, a landmark in theoretical science. Robert Mayer's treatment of the relationship between heat and work was accepted for publication only on the condition that he send in no more articles. In 1845, Joule proposed a theoretical (and today considered correct) explanation of the principles of the steam engine,

only to have it declared "mere nonsense" by the Royal Society of London. It was not until 1857 that scientists agreed upon what makes a steam engine work. In the field of chemistry the hypothesis of Avogadro, originally proposed in 1811 but largely ignored, was revived in 1860 and provided a significant turning point for the establishment of a sound chemical principle of atomic weights.

In art and music, as well as in a number of other activities, the acceptance of many new ideas has also been slow. The music of Bach was not uncovered until fifty years after his death, then to be recognized as some of the most inspiring music of all time. The work of such painters as El Greco and Vermeer remained relatively unrecognized for centuries before connoisseurs awakened to their value.

But today our attitude is changing. In many areas a thing must be *new* to have value. This is particularly true in the field of mechanics. The public clamors for new model cars each year, in spite of the fact that the engines may be essentially the same as in the previous models. We want new styles, new colors, new designs—and we are unusually willing to pay for it. Our houses, our radio and television sets, our clothes, our various household appliances, and even our vitamins and drugs must be the latest. It is indeed difficult to understand the occasional reluctance to accept change and improvement. For example, there is still bitter opposition to sanitation regulations and vaccination, even though these practices have been proved safe and successful.

What is the effect of these social attitudes upon physical science? This is indeed a complex problem, and several factors need consideration. With an increasing demand by society for more scientific and technological advances, we find today that the vast majority of our scientists, including the most competent, are supplying technical service to modern industry. These scientists, if they are to make continued advances in the development of fundamentally new concepts and principles, must be provided with an atmosphere that will foster this type of work. Instead, society still places too little value upon pure scientific study because it is still tardy in recognizing the value of pure scientific discovery. It may be that we have overemphasized the amenities of life in relation to our long-range needs, and society may be demanding that science cater to its desire for entertainment, comfort, beauty, and convenience to the detriment of the more permanent values inherent in increased knowledge.

30-6 The Social Responsibility of Scientists

At the present time, we occasionally hear the charge that scientists have cut themselves off from human or social concerns. The result, supposedly, is that science has produced things which society has not learned to use constructively. This charge has commonly been made in reference to the

invention of atomic bombs, toxic war gases, and other destructive devices, all of which have been made possible through the application of fundamental science.

Among several proposed solutions to this problem, two will be considered briefly. One proposal is to declare a moratorium on science for a length of time that will enable the other areas of human activity—social, political, economic, philosophical—to catch up. This proposal presumes that if the discovery of new scientific knowledge is stopped, these other areas will continue to advance at their own rates. But how could such a moratorium be initiated? To arrest the progress of science would necessitate stopping the scientists themselves, which would mean repressing the inherent, natural curiosity of man to inquire into the mysteries of nature. Because scientific curiosity in our best scientists is ineradicable, the scientists themselves would have to be destroyed, for how else could a government control their activities? Any committee or government agency set up to control scientists would soon be exercising similar control over economists, politicians, and possibly everyone else. The moratorium would have to be worldwide to be effective, but such policing action on a world scale has never been achieved in other areas, and it is difficult to conceive of it being accomplished in the scientific area alone.

Another proposal is to require that scientists direct their efforts toward constructive goals. This proposal, obviously advanced by people who are not acquainted with the ordinary operational procedures of science, would require that research and study be directed only toward beneficial goals. The history of science shows that a scientist gives his best service to humanity through fundamental study. The discoveries of X rays, penicillin, sulfa drugs, radio, television, and a host of other wonders have come as a direct result of fundamental, rather than applied, research. For instance, the modern "miracle" drugs were not discovered by the bedside physician but by scientists far removed from the hospital and sick bay.

Is it then possible to permit scientists to continue their fundamental studies only as long as the results are beneficial? Can we expect a scientist to foresee every evil use that someone might make of his discovery? We say emphatically "no." Let us trace the history of the atomic bomb and see if there was a point where the study of nuclear science could have been stopped. Rutherford's successful bombardment of the nucleus of the atom in 1911 was viewed at that time as a mere laboratory curiosity. Millions of alpha particles were required in order to achieve a single "hit"—such low efficiency could never be of practical significance. Chadwick's discovery of the neutron in 1932 raised no great reaction among scientists—it was impossible to predict at that time that the neutron would be the trigger for a chain reaction. Unless one were to depend upon the fiction writers, who always have been making fantastic predictions, there was no

dependable authority that could foresee the coming events. In 1934, Fermi bombarded uranium with alpha particles, but the results of his experiment were so unpredictable that he was unable to explain what had happened. In 1938–39, Hahn and Meitner in Germany sought to produce a trans-uranium element by the neutron bombardment of uranium, only to find that the uranium underwent fission. As happens so often, some of the most startling discoveries have been stumbled upon while searching for something else. By 1940, the achievements of nuclear science were classified as government secrets in order that the scientific findings could be directed toward military purposes. There was no stopping the further study of nuclear science at that point. In fact, during this period many scientists were pressed into the study of nuclear energy in order to make the construction of an atomic bomb possible.

It appears unreasonable, then, to expect scientists to be able to do more than anybody else, for scientists are ordinary human beings whose responsibility to society is no greater than that of other educated people. Society as a whole is responsible for using the findings of science wisely. To make an unsocial application of a scientific discovery is not unscientific, it is a misuse of information. It is conceivable that almost every scientific discovery has some good in it, and some bad. Whether we use the good or the bad depends upon society, and prior to the actual scientific discovery it is humanly impossible to predict which of these two uses society will choose.

Index

A

Aberration, 121
Absolute humidity, 459
Absolute magnitude, 545
Absolute zero, 397
Abundance of the elements, 248
Acceleration, 35, 78
Acceleration due to gravity, 41
Accumulation mountains, 108
Acetic acid, 327
Acid, 290, 324
Acid salts, 295
Activation energy, 337
Adams, John C., 125
Adenosine triphosphate, 358
Adhesive forces, 410
Adiabatic change, 463
ADP, 359, 364
Adsorbed molecules, 423
Advancing coast lines, 98
Age of the earth, 518, 519
Ages of the universe, 554
Airplane, 186
Alchemists, 4
Alcohol, 320, 324
Aldebaran, 545
Alluvial fan, 55
Alpha particle, 150, 225, 505
Alternating current, 171
Aluminum, 270, 380
Amide, 328
Amine, 327
Amino acid, 357
Ammeter, 161
Ammonia, 316, 345
Ammonia clock, 317
Amorphous solids, 430
Ampere, 161
Amplitude, 182
Amplitude modulation, 188
Anaerobic process, 360
Analytical balance, 19
Angstrom unit, 16
Anhydrite, 274
Anhydrous salts, 295
Anion, 376
Annual parallax, 116
Anode, 376
Anthracite, 107, 369
Anticline, 99
Ape man of Java, 26
Appalachian Mountains, 26, 100
Archimedes' principle, 4, 66
Arctic air mass, 465
Arcturus, 545
Area, 17
Arrhenius, Svante, 285
Aristotle, 40
Artificial transmutation, 509
Asbestos, 314
Aston, F. W., 211
Astronomy, 7, 588
Atlantic Ocean, 393
Atmosphere, 24, 70, 468
Atom, 4, 18, 148, 203, 226
Atomic nucleus, 251
Atomic number, 238
Atomic radii, 263
Atomic structure, 224
Atomic theory, 209
Atomic weight, 215
ATP, 359, 364
Aureomycin, 367
Aurora Borealis, 153, 168
Autotrophic organism, 352
Availability of useful energy from heat, 572
Average velocity, 37
Avogadro, Amedeo, 309, 407
Avogadro's law, 213
Avogadro's number, 221, 266

B

Babylonians, 4
Bacon, Roger, 4
Baking soda, 335
Barite, 66
Basalt, 105
Bases, 292
Basic salt, 295
Basin, 98
Batholith, 483
Beat, 194
Bequerel, Henri, 10, 149
Benzene, 322
Bergschrund, 61
Berzelius, 5, 208
Beta rays, 150, 225, 505
Betelgeuse, 545
Bethe, Hans, 538
Bevatron, 512
Binding energy, 508
Biochemistry, 351
Biogeology, 351
Biophysics, 351

Bituminous coal, 107, 369
Blast furnace, 281
Bohr, Niels, 227
Boyle, Robert, 5, 204, 244
Bradley, James, 122
Bragg, Sir William, 218
Bridgmann, P. W., 10
Bromine, 380
Bronsted, 292
Brown coal, 369
Brown, Robert, 418
Brownian movement, 419
Btu, 391
Bunsen, 5
Butylene, 319
Butyric acid, 327

C

Calcium, 380
Caldera, 489
Calendars, 3
Calcite, 299
Calorie, 391
Calvin, Melvin, 363
Camera, 194
Cannizzaro, Stanislao, 214
Capacitor, 152
Capillarity, 412
Carbohydrates, 353, 364
Carbon, 309, 517
Carbon black, 317, 380
Carbon compounds, 315
Carbon dating, 517
Cascade Mountains, 56
Casein, 425
Cassius' purple, 424
Catalyst, 325, 338
Cathode, 376
Cathode rays, 147
Cation, 376
Catskill Mountains, 108
Cause and effect, 595
Cavendish, Henry, 5, 46, 205
Cell structure, 585
Cellophane, 355
Celluloid, 354
Cellulose, 354
Celsius thermometric scale, 390
Cenozoic era, 26
Centauri, 545
Central Valley of California, 67
Centrifugal force, 88
Centripetal force, 87
Cepheid variables, 547
Chadwick, 151
Chain reaction, 337, 531
Chalcedony, 303
Chamberlain, T. C., 136
Charge on an electron, 148
Charles, Jacques, 398
Charles' law, 398
Chemical energy, 332
Chemical equations, 215
Chemical equilibrium, 341
Chemical erosion, 299
Chemical synthesis, 585
Chemistry, 6
Chemotherapy, 366
Chert, 107
Chilean earthquake, 64
Chlorine, 336, 380
Chloromycetin, 367
Chlorophyll, 275
Chromosphere, 452
Cinder cone, 487
Circular motion, 87
Circulation of the oceans, 472
Clays, 315
Clock, 23
Cloud chamber, 514
Coagulation of colloids, 424
Coal, 57, 396, 561
Cobalt-60, 533
Coenzymes, 359
Cohesive forces, 410
Cold fronts, 466
Color, 193, 195
Colorado River, 56
Colorado Rockies, 109
Colloids, 421
Combination of force, 84
Comets, 133
Compass, 164
Composite cone, 488
Composition of dry air, 249
Composition of the earth's crust, 250
Composition of the oceans, 249
Compound, 205, 208
Compton, Dr. Karl T., 387
Conduction, 399
Conductors, 145, 257
Conglomerates, 58
Conservation of energy, 92, 565
Conservation of heat energy, 567
Conservation of mass, 205
Conservation of mechanical energy, 564
Constellations, 542
Continental air mass, 466
Continental glacier, 61, 98
Convection, 399
Convection cell theory, 110
Convergent lenses, 194
Coordinate valence, 271
Coordination number, 274
Copernican theory, 119
Copernicus, Nikolaus, 4, 118
Copolymer, 321
Copper, 374
Cordilleran geosyncline, 102
Core, 105, 199
Coriolis effect, 468
Corona, 452
Corrosion, 281, 381
Cosmic particles, 63
Cottonseed oil, 355
Coulomb, 161
Coulomb, Charles A., 145
Coulomb's law, 145, 271
Covalence, 270
Covalent bond, 308
Crab Nebula, 549
Cracking, 319
Crescent phase, 132
Critical point, 444
Critical pressure, 444
Critical temperature, 444
Cro-magnon man, 27
Crookes, Sir William, 146
Crookes tube, 147
Crust, 105
Crystalline solid, 430
Curie, Marie Sklodowska, 225, 505
Cyclotron, 511

D

Dalton, John, 5, 207
Dalton's atomic theory, 207
daVinci, Leonardo, 594
Davy, Humphrey, 5
deBroglie, Louis, 234
Deductive reasoning, 8
Delta, 55
Democritus, 4, 203
Density, 20
Density of barite, 66
Density of the earth's crust, 47
Density of galena, 66
Density of limestone, 65
Density of sandstone, 65
Descriptive science, 3
Deuteron, 510
Devonian period, 26, 98
Dew point, 460
Dewar, Sir James, 404
Diamonds, 309
Diastase, 359
Diastrophism, 108
Diffusion, 419
Diphosphopyridine nucleotide, 363
Dipole, 276
Direct current, 171
Distribution of molecular velocities, 416
Direct proportion, 44
Distance, 14
Distillation, 319
Divergent lenses, 194
Dome, 98
Doppler principle, 182, 547
Downslope movements, 63
DPNH, 363

Dry adiabatic lapse rate, 463
Dry cell, 376
Dry ice, 443
Ductility, 257, 432
Dynamo, 169
Dyne, 28

E

Earth, 130, 199
Earthquake, 103, 199, 479
Earth's density, 47
Earth's equator, 65
Earth's heat, 479
Earth's magnetic field, 167
Earth's orbit, 63
Eclipse, 132
Eclipsing variable, 547
Efficiency of heat engines, 570
Efficiency of the modern car, 571
Egyptians, 4
Einstein, Albert, 6, 230
Einstein effect, 582
Elastic limit, 431
Electric current, 144
Electric well logging, 381
Electricity, 160
Electricity from magnetism, 169
Electrochemical cells, 376
Electrical energy, 27, 332
Electrochemical energy, 373
Electrodes, 375
Electrolysis, 263, 378
Electrolyte, 284, 376
Electromagnet, 169
Electromagnetic wave, 187
Electron affinity, 267
Electronegativities, 268
Electrons, 143, 150, 224
Electroscope, 144
Electrostatics, 142, 165
Electrovalence, 269
Element, 204, 244
Ellipse, 127
Emitter junction, 384
Empedocles, 4
Emulsion, 425
Endothermal energy, 333
Energy, 13, 161, 332
Energy in natural processes, 573
Energy levels, 235, 259
English system, 15
Entropy, 573
Enzymes, 358, 585
Epicenter, 104
Equatorial air mass, 465
Equilibrium, 85
Equinox, 23, 131
Equivalents, 265
Eras, 25
Erg, 28
Erosion, 24, 53, 55, 300, 302
Erosional mountains, 107
Escape velocity, 453
Ester, 324
Ethyl alcohol, 320
Ethylene, 319
Euclid, 4
Euclidean geometry, 17
Evaporation, 409
Evolution of the universe, 553
Exosphere, 459
Exothermal reactions, 333
Expansion due to heat, 396
Exponents, 14
Extrusive rocks, 302
Eye, 194

F

Fahrenheit temperature scales, 390
Falling bodies, 40
Faraday, Michael, 5, 265
Faraday's law, 265
Farsighted eyes, 194
Fats, 355
Fault, 102
Feldspar, 302, 348
Fermentation, 358
Ferromagnesium, 303
First law of thermodynamics, 568
Fissure flow, 489
Fleming, Sir Alexander, 7, 360
Flood plain, 56
Fluorine, 380
Fog tracks, 148
Folds, 99
Foliated rock, 107
Force, 28, 48, 76
Forces in equilibrium, 85
Forces of action and reaction, 83
Formic acid, 327
Formula, 208, 264
Fossil, 26, 369
Fossil fuels, 396
Fossilization, 368
Foucault pendulum, 121
Frank, Alberta, 64
Franklin, Benjamin, 5, 143, 160, 492
Free energy, 572
Free radical theory, 336
Freeze drying, 444
Freezing, 278
Freezing point, 285, 442
French Academy of Science, 15
Frequency, 182
Frequency modulation, 188
Friction, 39, 87
Fructose, 353
Fuel cell, 376
Fumaroles, 493
Functional groups, 325

G

Galaxy, 18, 550
Galena, 66
Galileo, 5, 31, 40, 45
Gamma rays, 149, 225, 505
Gamow, George, 555
Garnet, 313
Gases, 414, 434
Gasoline, 318
Gay-Lussac, Joseph L., 213, 398
Geiger-Muller counter, 414
Gels, 425
Generator, 169
Geologic ages, 518
Geologic events, 484
Geology, 6
Geosyncline, 101, 111
Geysers, 496
Giant molecule, 310
Gibbous phase, 132
Glacial ages, 457
Glaciers, 59
Glacier National Park, 102
Glass, 3, 315
Glucose, 353
Glycine, 357
Gram, 19
Gram molecular weight, 220
Grand Canyon in Arizona, 24, 56
Granite, 105, 314, 481
Gravitational constant, 46
Gravity, 31, 165
Gravity in natural processes, 53
Graphite, 309, 369
Graphs, 447
Great Ice Age, 61
Great Salt Lake, Utah, 59
Greeks, 4
Greenhouse effect, 454
Greenland, 61
Greenwich civil time, 22
Ground, 145
Gulf Stream, 393, 473
Guncotton, 354
Gypsum, 58, 274, 299

H

Half-life, 25, 506
Halley, Edmund, 134
Halogens, 308
Hanging valleys, 61
Hard water, 300
Harmonics, 184
Heat, 390
Heat capacity, 62
Heat conductivity of some common substances, 401
Heat constants of certain materials, 438
Heat energy, 322, 395

Heat engines, 569
Heat in biological processes, 497
Heat of fusion, 436
Heat of vaporization, 438
Heliocentric theory, 118
Helium, 508
Helium clock, 519
Hemoglobin, 275
Henry, Joseph, 5
Hertz, Heinrich, 188
Heterotrophic organism, 352
Hexachlorobenzene, 323
High-temperature research, 584
Himalayas, 66, 100
Historical theories of atoms, 203
History of science, 3
Horsepower, 54, 82
Horsepower-hours, 54, 64
Hot springs, 496
Huygens, 5
Hydrate, 295
Hydrocarbon, 317
Hydrogen, 205, 229, 270, 508, 534
Hydrogen bond, 277
Hydrolysis, 298
Hydronium ion, 292
Hyperbola, 44
Hypothesis, 9, 595

I

Ice, 177
Iceland, 61
Impulse, 80
Induced charges, 152
Induction, 151
Inductive reasoning, 8
Inertia, 77
Inner part of the earth, 47
Inorganic compounds, 312
Insolation, 456
Insulators, 145, 257, 399
Insulin, 357
Interference, 193, 218
Intermolecular forces, 409
Intrusive formations, 302, 482
Inverse proportions, 44
Ion, 269, 287
Ion exchange, 348
Ion product of water, 297
Ionization potential, 231, 267, 373
Ionosphere, 459
Isomer, 318
Isoprene, 321
Isostasy, 66, 99
Isothermal process, 445
Isotopes, 212, 509

J

Jeans, Sir James, 136
Jeffreys, Harold, 136
Joint, 102
Joule, 81
Jupiter, 71, 89, 123
Jurassic period, 26

K

Karst, 302
Kant, Immanual, 135
Katabatic winds, 72
Kekule, August, 10, 322
Kelvin, Lord, 7
Kelvin temperature scale, 398
Kelvin "water dropper," 152
Kennelly-Heavyside layer, 190
Kepler, Johannes, 5, 127
Kepler's law, 127
Kerosene, 318
Kilowatt-hours, 163
Kinetic energy, 27, 89
Kinetic-molecular theory, 408, 426
Krakatoa, 492
Krypton, 15

L

Laccolith, 482
Lampblack, 380
Landscape, 55
Landslides, 63
Lapilli, 492
Laplace, Pierre-Simon, 135
Lapse rate, 463
Lateral moraine, 61
Lava tube, 481, 490, 494
Lavoisier, Antoine, 5, 205
Law, 9
Law of combining volumes, 212
Law of definite proportions, 206
Law of diffraction, 218
Law of gravitation, 48
Law of mass action, 340
Law of multiple proportions, 207
Law of original horizontality, 97
Law of universal gravitation, 46
Lead, 25
Lead clock, 519
Lead storage cell, 377
Le Chatelier, Henri, 343
Lecithin, 356
Left-hand rule, 169
Length, 14
Lenses, 194
Leucippus, 4, 203
Lewis, G. N., 293
Lewis overthrust, 102
Libby, W. F., 517
Lift, 87
Light, 189, 193
Lightning, 153
Light-year, 16, 544
Limestone, 58, 107, 299, 347
Line spectrum, 196
Lipids, 355
Liquids, 433
Lisbon, Portugal, 103
Liter, 17, 19
Lithium, 380
Lithosphere, 249
Lode, 247
Longitude, 22
Longitudinal waves, 104, 180, 184
Long-range forces, 585
Lowell, Percival, 126
Low-temperature research, 584

M

Magellanic Clouds, 551
Magma, 481
Magnesium, 270, 380
Magnetism, 164
Magnetism from electricity, 168
Magnitude of stars, 545
Malleability, 257, 432
Mantle, 105
Marble, 107
Marconi, Guglielmo, 188
Mares, 131
Maritime air mass, 466
Mars, 71, 123
Mass, 13, 18, 47
Mass and energy, 566
Mass of earth, 71
Mass of moon, 68
Mass spectrograph, 211
Mathematics, 7, 38, 44
Mature valley, 55
Maxwell, James Clerk, 187
McKenzie River, 56
Measurement of radioactivity, 513
Measurement of temperature, 389
Measurement of time, 517
Mechanical energy from electrical energy, 170
Medial moraine, 61
Mendeleev, Dmitri, 253
Mercury, 123
Meridian, 22
Meson, 504
Mesozoic era, 26
Metallic conductors, 160
Metals, 3, 145, 257, 268
Metamorphic rock, 483
Metamorphism, 106
Meteorites, 135
Meteors, 133, 135
Meter, 15
Methane, 317, 337
Methyl alcohol, 325
Metric units, 15, 19, 20

Micron, 16
Milky Way, 550
Millikan, Robert, 231
Minerals, 364
Mira stars, 548
Mississippi River, 56
Moderator, 527
Moist adiabatic lapse rate, 464
Mole, 210, 220
Molecular quantities, 408
Moment, 86
Momentum, 79, 89, 234
Montmorillonite, 303
Moon, 22, 24, 46, 49, 68, 131
Moraines, 61
Moseley, H. G. J., 239
Moulton, F. R., 136
Mountains, 107
Mt. Etna, 493
Mt. Everest, 66
Mt. Lassen, 495

N

Naming of compounds, 217
Natural gas, 317, 396, 561
N-type semiconduction, 383
Neap tides, 70
Nearsighted eyes, 194
Nebula, 549
Negative charge, 142, 160
Neon, 211
Neptune, 124
Neptunium, 532
Neutralization, 293
Neutron, 150, 224, 570, 526
New moon, 132
Newton, 48, 78
Newton, Isaac, 5, 45, 204
Newton meter, 81
Newton's first law, 77
Newton's law of gravitation, 65
Newton's second law, 78
Newton's third law, 82
Nitrogen, 205
Nodes, 180
Nonmetals, 257, 268
Normal distribution, 417
Normal fault, 102
North Pole, 40, 42, 131
North Star, 16
Nova, 548
Nuclear atom, 504
Nuclear energy, 523
Nuclear fission, 528
Nuclear fusion, 534
Nucleus, 226
Nuee ardente, 493
Nylon, 328

O

Observation in physical science, 5, 93
Obsidian, 494
Occurrence of the elements, 246
Ocean, 26
Oceanography, 590
Oceans and the weather, 474
Oersted, Hans Christian, 168
Ohm, 162
Ohm's law, 162
Old age valley, 55
Oleic acid, 327
Olivines, 313
Opal, 303
Orbitals, 235
Orchestra A, 186
Ores, 248
Organic acids, 326
Organic substances, 312
Osmosis, 413
Osmotic pressure, 414
Oxides, 290
Oxygen, 205, 290
Ozark Plateau, 108
Ozonosphere, 455

P

Paleozic era, 26
Parabola, 44
Paracelsus, 4
Paradichlorobenzene, 323
Parallax, 120
Parallel circuit, 174
Paricutin, 487
Peat, 369
Pelean cloud, 493
Penicillin, 7, 360
Periods, 254
Periodic classification of the elements, 252
Periodic law, 254
Periodic table of the elements, 254
Perrin, Jean, 419
Petroleum, 319, 396, 561, 571
pH of acids, bases and neutral systems, 297
pH of water, 296
Phenylalanine, 357
Philosopher's stone, 4
Phospholipide, 356
Photochemical laws, 362
Photochemical process, 337
Photoelectric effect, 231
Photon, 225, 230, 361
Photosphere, 452
Photosynthesis, 361
Physical science defined, 2
Physics, 6
Pigment color, 196
Planck, Max, 6, 230
Planetoids, 133
Planets, 4, 116, 123
Plasma, 589
Pleistocene, 26
Pleochroic halos, 520
Pliny the Elder, 4
Pluto, 123
Plutonium, 532
Polar air mass, 465
Polar front, 471
Polonium, 149
Polyethylene, 320
Polymers, 320
Positive charge, 142, 160
Positivism, 593
Positron, 505
Potential difference, 161
Potential energy, 27, 89, 91, 373
Poundal, 49
Power, 82
Precambrian, 26
Precipitation, 57
Pressure, 71
Pressure effect on materials, 445
Prevailing westerlies, 471
Priestley, Joseph, 5, 204, 290
Primary cell, 377
Primary coil, 172
Primary rainbow, 197
Primary wave, 199
Principia, 45
Principle of superposition, 484
Principle of uniformitarianism, 484
Principal quantum numbers, 235
Prism, 195
Probability, 417, 595
Prominences, 452
Proportion, constant of, 32
Proteins, 356, 585
Proton, 150, 224, 510
Proton acceptor, 292
Proton donor, 292
Proust, Joseph, 206
Ptolemaic theory, 117
Ptolemy, 4, 117
p-type semiconduction, 383
Pumice, 494
P-waves, 104
Pyroxenes, 314
Pyroxylin, 354
Pythagoras, 4

Q

Quanta, 234
Quartz, 23, 107, 303, 481
Quaternary period, 26
Quinine, 366

R

Radar, 189, 190
Radial motion, 547
Radiant energy, 332
Radiation, 399
Radicals, 271

Radio, 188
Radio stars, 550
Radioactive clock, 25
Radioactivity, 149, 480, 505
Radioastronomy, 588
Radioautogram, 516
Radium, 149
Rainbow, 196
Rainstorm, 54
Ramsay, William, 419
Rate of chemical reaction, 340
Rayon, 355
Reactants, 216
"Reaction complex," 339
Reaction mechanisms, 334, 346
Reactivity, 245
Reactor, 531
Real image, 190
Realism, 593
Receding coast lines, 98
Recessional moraines, 61
Red-shift, 552
Reduction, 279
Relative humidity, 459
Reflection, 189
Reflection shooting, 191
Refraction, 191, 192, 199
Relativity, 581
Resonance, 185, 323
Respiration, 360
Reserpine, 367
Resultant force, 85
Retrograde motion, 119
Reversible reaction, 342
Richter, Jeremias, 206
Rigel, 545
Rocks, 24
Rocky Mountains, 26, 109
Roentgen, Wilhelm, 197
Rubber, 321
Rumford, Count, 395
Rust, 382
Rutherford, 6

S

Salt, 57, 58, 209, 294
San Andreas rift, 103
San Francisco earthquake, 103
Sandstone, 57, 58, 107, 302
Satellites, 133
Saturated solution, 288
Saturation humidity, 459
Saturn, 124
Scalars, 34
Schist, 107
Scientific law, 507
Scientific method, 2, 8
Scientific theory, 121
Scintillometer, 515
Second, 23
Second law of thermodynamics, 572
Secondary cell, 377
Secondary coil, 172
Secondary rainbow, 197
Secondary waves, 199
Sedimentary materials, 24
Sedimentary rock, 57
Seismograph, 104, 199
Semiconductor, 382
Separation of isotopes, 536
Series circuit, 173
Shale, 57, 58, 107
Shield volcano, 488
Siderial year, 23
Sierra Nevadas, 102, 109
Silicates, 303
Silicon, 311
Sill, 482
Sinks, 302
Sirius, 545
Slate, 107
Smog, 63
Soap, 3, 300
Social attitudes toward physical science, 599
Social responsibility of scientists, 600
Sodium chloride, 220, 264, 379
Soddy, 6
Solar battery, 562
Solar energy, 537, 588
Solar furnace, 562
Solar radiation, 479
Solar system, 63, 115, 453
Solar temperature, 129
Solids, 429
Solstices, 23
Solubility, 59
Solute, 288
Solution, 288, 421
Solvent, 288, 421
Sonic sounding, 191
Sound waves, 184
Sources of energy, 560
South Pole, 40, 72, 131
Space, 13, 578, 586
Specific gravity, 21
Specific heat, 391
Specific heats of some common substances, 393
Spectroscope, 194
Spectrum, 195
Speed, 32, 88
Spitzbergen, 61
Spring tides, 70
Stalactite, 301
Stalagmite, 301
Standard electrochemical potentials, 375
Standing waves, 180
Starch, 354
Stars, 22, 544
Stearic acid, 327
Storage battery, 377
Strassfurt deposits, 59
Stratosphere, 459
Streptomycin, 367
Strike-slip fault, 102
Strong acids, 292
Structural color, 196
Sublimation, 443
Sucrose, 353
Sugar, 209
Sulfanilamide, 367
Sulfur, 309
Sun, 22, 129, 452
Sunspots, 452
Superior air mass, 466
Supernovae, 549
Supersaturated solution, 288
Surface area, 423
Surface catalyst, 339
Surface tension, 410
Suspension, 421
S-waves, 104
Swiss Alps, 100
Symbols, 244
Synchrotron, 513
Syncline, 99

T

Taylor, Lloyd W., 387
Telescope, 194
Temperature, 390
Temperature effect on materials, 445
Tensile strength, 257
Terminal moraine, 61
Terramycin, 367
Tetrahedral arrangement, 313
Textiles, 3
Theories on the evolution of the solar system, 135
Theories of mountain formation, 109
Theory, 9
Thermal neutrons, 527
Thermodynamics, 568
Thermometer, 389
Thermonuclear reaction, 535
Thompson, Benjamin, 395
Thomson, J. J., 147, 211
Thrust fault, 102
Tides, 24, 68
Time, 13
Tin, 311
Tombaugh, Clude, 126
Torque, 86
Tracer element, 516
Tranquilizers, 368
Transformer, 172
Transpiration, 440
Transverse waves, 104, 179, 184
Trinitrotoluene, 324
Tropic of Capricorn, 131
Tropical air mass, 465
Triple point, 443

Tropopause, 459
Troposphere, 459
Tycho Brahe, 5, 127
Tyndall effect, 422

U

Ultraviolet, 456
Unbalanced forces, 87
Unfoliated rocks, 107
Unsaturated hydrocarbons, 320
Unsaturated solution, 288
Upper atmosphere, 586
Uranium, 25, 149, 507, 529
Uranus, 123
Urey, Harold C., 137
U-shaped valley, 60
Utilization of nuclear energy, 531

V

Vacuum bottle, 404
Valence, 263
Valley glacier, 60
Valley of Death, 72
Van Allen belt, 587
Van de Graaf generator, 513
Vapor pressure, 439
Variable stars, 547
Variation in gravity, 65
Vectors, 34
Vega, 545
Vein, 102, 247
Velocity, 32, 88
Velocity of gas molecules, 415
Velocity of sound, 182
Venus, 71, 123
Vesuvius, 493
Vinegar, 3
Virtual image, 190
Vitamins, 364
Volcanic activity, 491
Volcanic ash, 303
Volcanoes, 63
Volt, 161, 373
Volume, 17
von Laue, Max, 218
von Weizacker, C. F., 137, 538
V-shaped valley, 56
Vulcanization, 321

W

Warm front, 467
Warps, 97
Water, 276, 399, 434
Water gas, 325
Water glass, 313
Water in natural processes, 299
Water in the air, 459
Water wave, 180
Watt, 82, 163
Watt-hour, 163
Wave length, 182
Wave mechanics, 235
Wave motion, 179
Weak acids, 292
Weathering, 347
Weight, 18, 47, 78
Wet- and dry-bulb thermometer data, 462
Whistlers, 587
Wilson, C. T. R., 514
Winter solstice, 130
Wood, 369
Wood alcohol, 325
Work, 27, 81, 90
Work hardening, 430
World's fossil fuel supply, 561

X

X ray, 5, 197, 240
X-ray diffraction, 218

Y

Yellowstone Park, 64
Yosemite Valley in California, 61
Youthful valley, 55
Yukawa, H., 504

Z

Zinc, 374
Zymase, 358